Analysis of post-transcriptional gene expression modulated by mRNA stability and RNA-binding proteins in human cells

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Jennifer V. Gerbracht

aus Mainz

Veröffentlichung: Köln, 2020

Berichterstatter: Prof. Dr. Niels Gehring

(Gutachter)

Prof. Dr. Kay Hofmann

Prof. Dr. Elmar Wahle

Tag der mündlichen Prüfung: 06.02.2020

CONTENTS

1 INTRODUCTION

In this thesis, three different aspects of RNA biology are investigated, all of which depend on the interplay between RNA and proteins. The first part examines the processing of ribosomal RNAs (rRNAs) which are the catalytic part of mature ribosomes and thus essential for protein biosynthesis. The genetic information is transmitted to the ribosomes in form of messenger RNA (mRNA). In order to be stably expressed, exported and translated, mRNAs are bound by RNA-binding proteins, thereby forming messenger ribonucleoproteins (mRNPs). In metazoans, the multimeric exon-junction complex (EJC) is an integral component of mRNPs. As a mark on spliced mRNAs, EJCs constructively influence gene expression. The assembly and regulatory functions of the EJC and its peripheral interacting factors are examined as the second part of this thesis. Finally, mRNAs undergo turnover to allow the cell to adapt its gene expression landscape to current demands. In the third part of the thesis, mRNA degradation pathways are described and a new method to study the contribution of different decay pathways is presented. All processes in this thesis are described for human cells, except when mentioned otherwise.

1.1 Regulation of gene expression on the level of RNA

Gene expression is mediated and regulated by a multitude of RNA species in the cell. Quantitatively, mRNAs as the transmitter of genetic information represent only about two percent of total cellular RNAs (Frith *et al.*, 2005). The bulk of the total RNA pool is made up of non-coding RNAs (Mattick, 2001). These include transfer RNAs (tRNAs) and rRNAs as important members of the protein synthesis machinery. Additionally, a vast landscape of small non-coding RNAs has been described which are involved in regulatory processes such as splicing (small nuclear RNAs, snRNAs), ribosomal biogenesis (small nucleolar RNAs, snoRNAs) and RNA silencing (small interfering RNAs, siRNAs and microRNAs, miRNAs). In gene expression steps such as splicing and translation, RNA fulfills a catalytic role, an ability which for a long time had been attributed exclusively to proteins (Cech, 2009).

A diverse set-up of RNA polymerases and RNA processing pathways underlie the production of these RNAs in the nucleus. This chapter describes the synthesis and processing of rRNAs and mRNAs.

1.1.1 Ribosomal RNA maturation and ribosomal biogenesis

Ribosomes as the sites of protein biosynthesis play an essential part in the gene expression apparatus. The eukaryotic 80S ribosome comprises two subunits, the 40S small subunit (SSU) and the 60S large subunit (LSU) (Khatter *et al.*, 2015). Four rRNA species and more than 80 proteins must be produced, processed and assembled to form mature ribosomes. This highly complex process requires spatial and temporal coordination across different compartments of the cell as well as the assistance of more than 200 ribosome biogenesis factors (Figure 1A) (Tafforeau *et al.*, 2013).

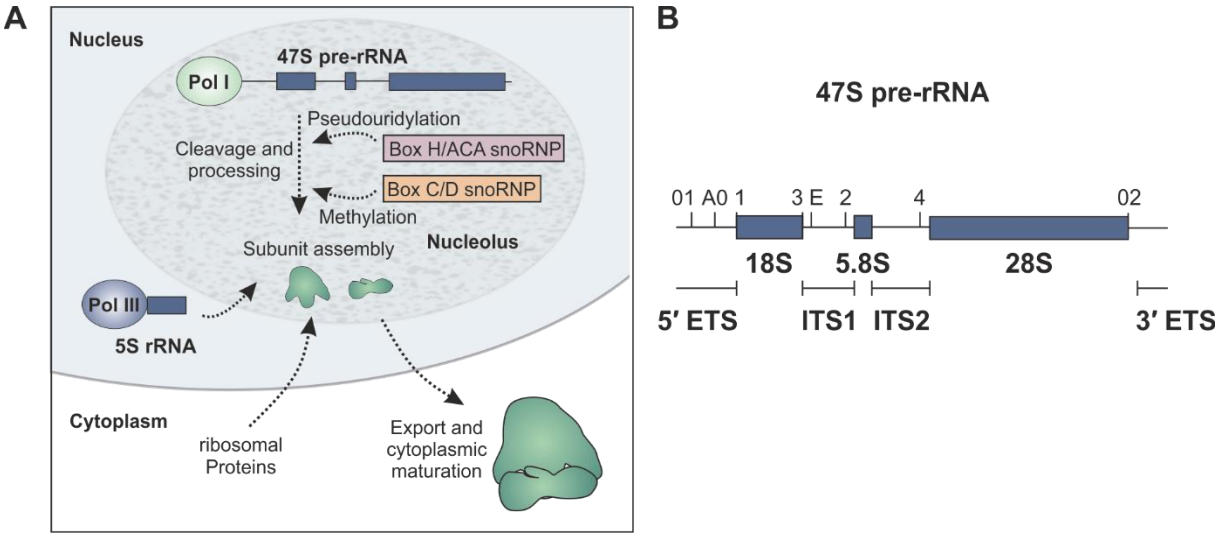


Figure 1: Overview of ribosomal biogenesis. A*: The nucleolus is the hub of ribosomal biogenesis within the nucleus. It forms around clusters of ribosomal RNA (rRNA) genes which are the sites of rRNA transcription by RNA polymerase I. Here, the 47S pre-mRNA is cleaved, modified and associates with ribosomal proteins. The 5S rRNA is transcribed by RNA Polymerase III and travels to the nucleolus to be assembled into the large ribosomal subunit. The final maturation of the small and large ribosomal subunits occurs after export into the cytoplasm.* ***B****: Schematic depiction of the 47S pre-rRNA that harbors the 18S, 5.8S and 28S rRNAs. These rRNAs are produced by cleaving the precursor at the indicated endonucleolytic cleavage sites. The external and internal transcribed spacers (ETS and ITS) are removed by combined endo- and exonucleolytic cleavage.*

Ribosomal biogenesis begins in the nucleolus, which exists as a membrane-less subcompartment within the nucleus (McStay, 2016). Reminiscent of bacterial gene expression, the 47S rRNA precursor (pre-rRNA) carrying the 18S, 5.8S and 28S rRNAs is transcribed by RNA polymerase I as a polycistronic transcript (Figure 1B) (Henras *et al.*, 2015). The rRNAs are separated by internal transcribed spacers (ITS1 and 2) and encompassed by 5' and 3' external transcribed spacers (5'-ETS and 3'-ETS). An essential part of rRNA maturation is exonucleolytic trimming from the 5'- and 3'-end as well as endonucleolytic cleavage (endocleavage) steps during which the transcribed spacers are removed. The separation of

the rRNAs required for both ribosomal subunits occurs by cleavage within the ITS1 region. After final processing, the 18S rRNA is incorporated into the SSU, while the 5.8S and 28S rRNA are constituents of the LSU (Henras *et al.*, 2015).

Co-transcriptionally, the pre-rRNA associates with ribosomal proteins and ribosome biogenesis factors forming pre-ribosomes. In these large dynamic structures, rRNA folding and modification as well as the assembly of ribosomal proteins takes place (Kressler *et al.*, 2017).

The modification of rRNA residues is essential for proper rRNA folding and ribosome function (Sloan *et al.*, 2017). This process is carried out by ribonucleoprotein complexes formed by snoRNAs and associated proteins (snoRNPs). The two major types of modification are pseudouridylation catalyzed by box H/ACA and methylation by box C/D snoRNPs (Figure 1A) (Watkins & Bohnsack, 2012). These modifications occur predominantly in functionally important regions (Sloan *et al.*, 2017).

The genes encoding for ribosomal proteins are transcribed in the nucleus by RNA polymerase II, translated in the cytoplasm and the proteins reimported and transported to the nucleolus (Kressler *et al.*, 2017). The fourth rRNA species, the 5S rRNA transcribed by RNA polymerase III, also joins the pre-ribosomes in the nucleolus as a preassembled RNP (Ciganda & Williams, 2011). The fully assembled subunits are then transported out of the nucleus and undergo further maturation in the cytoplasm (Kressler *et al.*, 2017).

Ribosomal biogenesis is a process which is highly energy consuming for the cell (Warner, 1999). It requires the transcription of RNAs by all three RNA polymerases, the synthesis of the most abundant RNAs and proteins in the cell and the production of ribosome biogenesis factors. It is therefore a tightly controlled process and undergoes quality surveillance (Pena *et al.*, 2017). Furthermore, the nucleolus as a hub of ribosome biogenesis serves as a cellular stress sensor (Golomb *et al.*, 2014). The impairment of rRNA transcription, rRNA processing, and ribosome subunit assembly results in nucleolar stress, leading to the stabilization of the tumor suppressor p53 and subsequent cell cycle arrest (Bursac *et al.*, 2014).

1.1.2 Messenger RNA processing

Protein-coding genes are transcribed by RNA Polymerase II (Figure 2) (Cramer, 2019a). As soon as the first nucleotides of pre-mRNA emerges from RNA polymerase II during transcription elongation, maturation is initiated by distinct machineries that are spatially

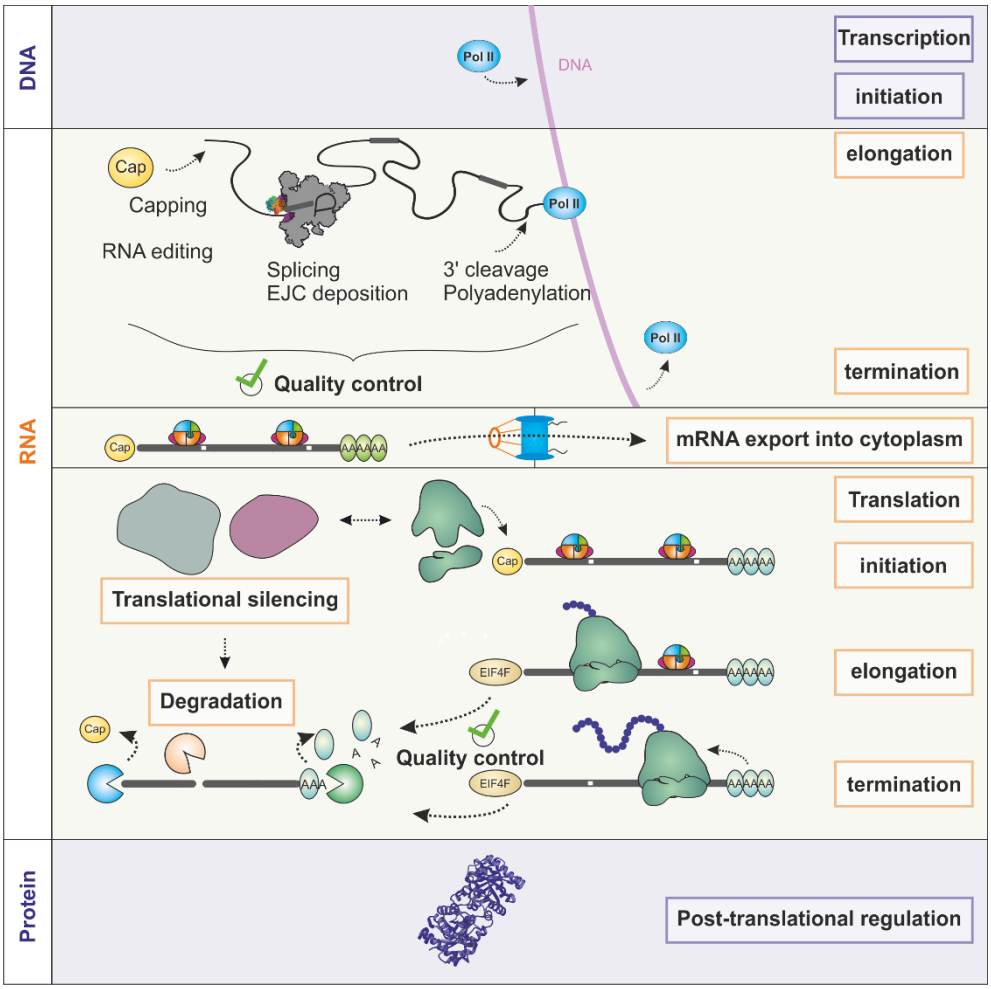


Figure 2: Steps of post-transcriptional gene regulation that affects messenger RNAs. *After transcription initiation, mRNA processing occurs co-transcriptionally. The phosphorylation of the RNA polymerase II C-terminal domain (CTD) and the rate of transcription affect the recruitment of the RNA processing machinery. Factors involved in capping, splicing, RNA modification and polyadenylation are present at the site of transcription. Erroneously transcribed or processed mRNAs are degraded in the nucleus. mRNA transcription is therefore directly coupled to quality control by mRNA processing and degradation machineries. For representational reasons, the depiction of messenger ribonucleoprotein particles (mRNPs) focuses on mRNA bound by exon junction complexes (EJCs). After transcription, mRNA-binding export factors mediate the export through the nuclear pore. In the cytoplasm, translation is initiated by interactions between the cap binding complex, translation initiation factors, and the small ribosomal subunit. Upon stress, mRNAs can be moved to membrane-less compartments such as P-bodies and stress granules and remain translationally silent. All transcripts undergo turnover by the degradation machinery. mRNA degradation is also directly coupled to translation during mRNA surveillance, e.g. in the nonsense-mediated mRNA decay (NMD) pathway which will be described in detail below. Following translation, gene expression is modulated on the protein level. This includes post-translational modifications that influence the function of proteins or alterations to the stability of a protein.*

and functionally linked. Central to this process is the C-terminal domain (CTD) of RNA polymerase II which consist of heptad repeats that are post-translationally modified (Harlen & Churchman, 2017). These modifications as a so-called CTD code change dynamically throughout the transcription process (Buratowski, 2003). They are crucial for the transcription steps and are required for 5' capping, splicing and 3' end processing (Harlen & Churchman, 2017). By coupling transcription to processing, certain checkpoints are established that protect the cell against the production of aberrant mRNAs (Peck *et al.*, 2019).

1.1.2.1 Capping

The first co-transcriptional modification is the capping of the mRNA which occurs after the transcription of 25-30 nucleotides (Ramanathan *et al.*, 2016). In three enzymatic steps, a 7-methylguanosine cap is added to the first nucleotide of the nascent transcript (Ghosh & Lima, 2010). The binding of the capping enzymes to the CTD phosphorylated at the Serine-5 position (Ser5) is required for efficient capping (McCracken *et al.*, 1997, Komarnitsky *et al.*, 2000, Schroeder *et al.*, 2000). The 5' cap is bound co-transcriptionally by a nuclear cap-binding complex (CBC). This complex directly interacts with components of the splicing machinery and thus links capping to splicing (Pabis *et al.*, 2013).

1.1.2.2 Splicing

The splicing reaction requires the assembly of five small nuclear ribonucleoproteins (snRNPs) U1, U2, U4, U5 and U6 as well as 85 non-snRNP proteins as the core spliceosome (Kastner *et al.*, 2019). The spliceosome catalyzes the excision of an intron in a two-step transesterification reaction. Recently published spliceosome cryo-electron microscopy (cryo-EM) structures have illuminated the compositional dynamics and structural rearrangements of the splicing reaction in yeast and humans (Kastner *et al.*, 2019). It is estimated that the majority of splicing reactions occur co-transcriptionally, but some introns are spliced out after transcription termination (Coulon *et al.*, 2014). Supporting this notion, the Ser5 phosphorylated CTD of RNA polymerase II was found to interact with components of catalytically active spliceosomes (Nojima *et al.*, 2018).

The U1 and U2 snRNAs are essential for splice site selection by base paring to splice site consensus sequences (Figure 3) (Papasaikas & Valcarcel, 2016). Since the genome contains a myriad of potential splice sequences, additional mechanisms exist to define exons and introns for correct splicing. In human cells, typically short exons (median length of ~ 130 nucleotides)

are separated by longer introns (median length of ~1.7 kilobases) (Piovesan *et al.*, 2016). It was proposed that splicing complexes would form across an exon (exon definition model) (Robberson *et al.*, 1990). Indeed, exon-covering complexes that contain all snRNPs have been identified (Schneider *et al.*, 2010). The sequences flanking potential splice sites can either enhance or inhibit exon inclusion by their interaction with splicing factors that can positively or negatively influence splice site selection (De Conti *et al.*, 2013). Exonic splicing enhancers (ESE) are commonly bound by proteins of the serine/arginine-rich family (SR) which promote spliceosomal complex formation (Graveley *et al.*, 2001). The repressing effect of exonic splicing silencing regions (ESS) is primarily mediated by proteins of the hnRNP family (Zhu *et al.*, 2001).

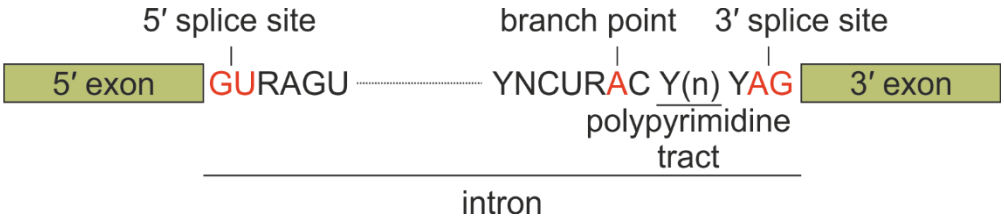


Figure 3: Conserved sequences at 5′ and 3′ splice sites in metazoans. *R: purine, Y: pyrimidine, N: nucleotide. The 5′ splice site, branch point and 3′ splice site are marked in red. According to Will & Luhrmann (2011).*

It is estimated that the vast majority of eukaryotic genes undergoes alternative splicing (Pan *et al.*, 2008). Alternative splicing increases the number of potential gene products that are encoded within a genomic region. Furthermore, it can have a regulatory function, e.g. by switching from a stable mRNA isoform to an isoform that is subject to mRNA decay (Nasif *et al.*, 2018). Alternative splicing can be regulated by RNA-binding proteins and changes depending on transcription elongation speed (Dujardin *et al.*, 2014, Fong *et al.*, 2014, Ule & Blencowe, 2019). Transcription by RNA polymerase II does not occur at a steady pace. Depending on transcription elongation factors, histone dynamics and DNA/RNA hybrid stability, RNA polymerase II can pause in the proximity of the promoter, backtrack and proceed at varying speeds (Saldi *et al.*, 2016, Core & Adelman, 2019, Cramer, 2019b). Changing the elongation rate potentially influences the selection of splice sites or RNA structures that determine the outcome of splicing (Saldi *et al.*, 2016). It was hypothesized that binding of competing splice factors is influenced by the duration of which splice sites are available, although this model only explains a subset of observed events (Fong *et al.*, 2014). The function of the transcription elongation rate in transcription termination is better described (see below).

1.1.2.3 3′ processing

The 3′ end processing machinery scans the pre-mRNA for the polyadenylation site, performs cleavage and adds a polyA tail (Kumar *et al.*, 2019). Since the Ser2 phosphorylation of the CTD is a prerequisite for the recruitment of cleavage/polyadenylation factors, 3′ end processing is also directly coupled to transcription (Ahn *et al.*, 2004). Premature 3′ cleavage at cryptic polyadenylation sites is prevented by the telescripting activity of the U1 snRNP in a complex with cleavage and polyadenylation factors (So *et al.*, 2019). There are two leading hypotheses on how transcription termination occurs. Firstly, a conformational or functional change of the transcription elongation complex (allosteric model) or secondly, degradation of the nascent transcript after 3′ cleavage by the 5′-3′ exonuclease XRN2 (torpedo model) (Porrua & Libri, 2015). Recently, a model termed sitting duck was proposed that connects these hypotheses: when the RNA Polymerase II elongation complex has proceeded downstream of the polyA site, the dephosphorylation of a transcription elongation factor slows down RNA Polymerase II (Cortazar *et al.*, 2019). This enables XRN2 to catch up with the decelerated elongation complex.

1.2 Functions of the EJC in the lifetime of an mRNA

The cellular fate of an mRNA is determined by RNA-binding proteins (Gehring *et al.*, 2017). A prime example is the EJC which is deposited on transcripts during splicing and stays bound until it is removed during translation (Dostie & Dreyfuss, 2002, Lejeune *et al.*, 2002). The complex serves as a mark of spliced mRNAs and acts as a binding platform for peripheral EJC-associated factors. Both EJC core components and peripheral interacting factors generally have a stimulatory effect on gene expression. Furthermore, the EJC has a well-characterized function in the nonsense-mediated mRNA decay (NMD) pathway.

1.2.1 Assembly of the EJC core during splicing

The assembly of the EJC starts with the recruitment of EIF4A3 by the spliceosome component CWC22 (Alexandrov *et al.*, 2012, Barbosa *et al.*, 2012, Steckelberg *et al.*, 2012, Steckelberg *et al.*, 2015). EIF4A3 directly binds to the mRNA in a sequence-independent manner by interacting with the sugar-phosphate backbone of the RNA (Chan *et al.*, 2004, Ferraiuolo *et al.*, 2004, Palacios *et al.*, 2004, Shibuya *et al.*, 2004, Andersen *et al.*, 2006, Bono *et al.*, 2006). It is a member of the DEAD-box family of ATP-dependent RNA helicases (Linder & Jankowsky,

2011). The binding of the heterodimer MAGOH-RBM8A locks EIF4A3 into its RNA-bound conformation by preventing the release of ATP hydrolysis products (Nielsen *et al.*, 2009). The mapping of the EJC position showed that eight nucleotides at a conserved position 20-24 nucleotides upstream of an exon-exon junction are protected from RNase H digestion by the EJC (Le Hir *et al.*, 2000). High-throughput analyses of RNA-interaction sites of EJC components confirmed the canonical position of the EJC upstream of exon-exon junctions (Sauliere *et al.*, 2012, Singh *et al.*, 2012).

It is possible to assemble the EJC *in vitro* by combining recombinant EJC components together with single stranded RNA and a non-hydrolysable analogue of ATP (Ballut *et al.*, 2005). These experiments have shown that the addition of the EJC interacting domain of CASC3 (called SELOR domain) enhances the stability of *in vitro* formed EJCs. Furthermore, early studies found that *in vivo* expressed EJC components form a tetrameric complex (Tange *et al.*, 2005). It was therefore concluded that CASC3 together with EIF4A3, RBM8A and MAGOH represent the essential EJC core. In the crystal structures of the EJC, EIF4A3, RBM8A and MAGOH as well as two fragments of SELOR are resolved (Figure 4) (Andersen *et al.*, 2006, Bono *et al.*, 2006). These structures revealed interaction sites between the factors and between the EJC and RNA. Unlike MAGOH-RBM8A, SELOR is also able to bind EIF4A3 in an open, RNA-unbound conformation (Bono *et al.*, 2006). By doing so, the ATPase activity of EIF4A3 is increased (Noble & Song, 2007).

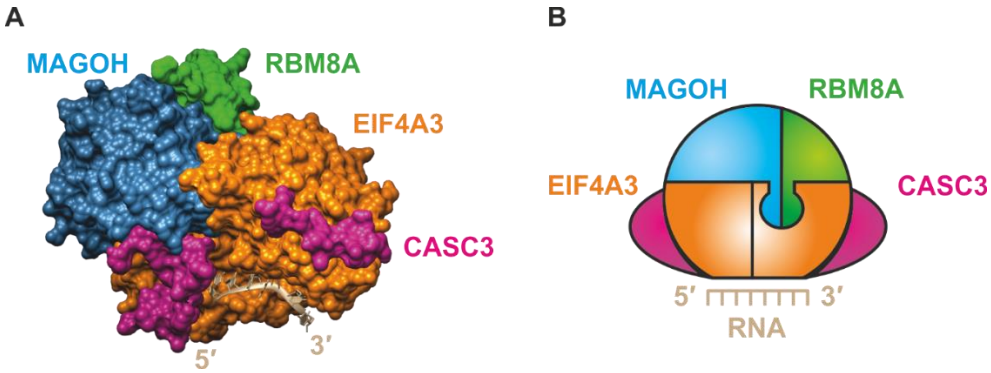


***Figure 4: The structure of the EJC. A**: The crystal structure of the EJC (pdb: 2J0S) rendered with Chimera v.1.14rc consisting of EIF4A3 (orange), MAGOH (blue), RBM8A (green) and CASC3 (magenta). Six nucleotides are bound by EIF4A3 and are shown in beige. **B**: Schematic depiction of the complex.*

Several observations contradict the concept of CASC3 as part of the tetrameric EJC core. CASC3 is a nucleocytoplasmic shuttling protein that is mainly located in the cytoplasm, whereas the other EJC core components are predominantly detected in the nucleus (Kataoka

et al., 2000, Le Hir *et al.*, 2001a, Degot *et al.*, 2002, Palacios *et al.*, 2004, Shibuya *et al.*, 2004). It was hypothesized that the low concentration of CASC3 in the nucleus could prevent the spontaneous assembly of the EJC and that it had to be actively recruited during splicing (Tange *et al.*, 2005). However, *in vitro* splicing experiments showed that the deposition of EIF4A3, MAGOH and RBM8A is strictly dependent on splicing, whereas CASC3 can associate with a preformed trimeric EJC (Gehring *et al.*, 2009a). Unlike the other three EJC core components, CASC3 was not identified in purified spliceosomal C complexes (Bessonov *et al.*, 2008, Agafonov *et al.*, 2011). Physiological studies of mice depleted of EJC components have also hinted at a differential function of CASC3: the EJC is essential for survival as was demonstrated by MAGOH (-/-) mice that die during an embryonic development stage (Silver *et al.*, 2010). Mice that are haploinsufficient for MAGOH or RBM8A display severe developmental and neurological defects (Silver *et al.*, 2010, Mao *et al.*, 2015). While CASC3 (-/-) mice are also not viable, developmental defects in CASC3 haploinsufficient mice are much milder compared to the other EJC components (Mao et al. 2017). For these reasons, the term “core EJC” will be subsequently used to describe the trimeric pre-EJC formed by EIF4A3, MAGOH and RBM8A.

1.2.2 Regulatory function in transcription and splicing

Since the deposition of the EJC is dependent on splicing itself, the finding that the correct splicing of the MAPK gene in *Drosophila* requires the core EJC components was intriguing. (Ashton-Beaucage *et al.*, 2010, Roignant & Treisman, 2010). The absence of the EJC results in the skipping of multiple exons resulting in impaired MAPK protein expression. While the fly homologue of CASC3 was not implicated for this function, depletion of the EJC interactor and the previously characterized splicing factor RNPS1 also caused alternative splicing of the MAPK mRNA. Transcriptome-wide, the splicing of long introns was specifically shown to be dependent on the EJC in *Drosophila*. RNPS1 forms together with ACIN1 and SAP18 the ASAP complex that interacts with the EJC (Schwerk *et al.*, 2003, Tange *et al.*, 2005). An alternative complex with PNN instead of ACIN1 can also be formed (Murachelli *et al.*, 2012). Components of the ASAP complex together with the EJC were found to prevent intron retention of the Piwi mRNA (Hayashi *et al.*, 2014, Malone *et al.*, 2014). It was proposed that the deposition of the EJC and subsequent recruitment of interacting splice factors could have an exon defining function necessary for the splicing of large neighboring introns (such as MAPK) or introns with weak splice consensus sequences (such as Piwi). Interestingly, for the Piwi mRNA the splicing

of downstream junctions affected intron retention upstream (Malone *et al.*, 2014). This argues against a strict 5'-3' hierarchy of splicing order for this particular transcript.

In *Drosophila*, the knockdown of EJC core components leads to a decrease of RNA polymerase II occupancy at the 5' end of transcripts and an increase on the gene body during transcription (Akhtar *et al.*, 2019). This reduction of promoter proximal pausing of RNA polymerase II is partly caused by increased phosphorylation of Ser2 in the CTD. In *Drosophila*, this process was proposed to be the cause of a subset of alternative splicing events observed upon EJC knockdown (Akhtar *et al.*, 2019).

The first analysis of transcriptome-wide EJC-dependent splicing in human cells yielded a high number of transcripts that were affected (Wang *et al.*, 2014b). Similarly to the observations made in *Drosophila*, it was reported that in human cells the depletion of EJC core components leads to an increase in the transcription elongation speed on selected genes, potentially affecting alternative splicing. In contrast to the studies performed in *Drosophila*, the trimeric EJC core factors as well as CASC3 affected EJC-dependent genes in the same manner.

1.2.3 Role in mRNA export

Mature mRNAs are transported out of the nucleus through the nuclear pore complex for their translation in the cytoplasm. Components of the transcription and export (TREX) complex are associated with the mRNA during its maturation (Strasser *et al.*, 2002). Export occurs when the nuclear export factor 1 (NXF1, also known as TAP) is recruited by TREX adaptor proteins. NXF1 forms a heterodimer with NXT1 (also known as p15) and mediates translocation through the nuclear pore (Heath *et al.*, 2016).

Splicing has been shown to promote mRNA export (Valencia *et al.*, 2008). Since the TREX components DDX39B (UAP56) and ALYREF were found to interact with the EJC it was suggested that the EJC stimulates export as a binding platform for mRNA export factors (Le Hir *et al.*, 2000, Le Hir *et al.*, 2001b). ALYREF recruitment was shown to be 5' cap dependent and that binding occurs in the first exon (Cheng *et al.*, 2006). A short linear motif in ALYREF is required for its interaction with EIF4A3 and the CBC (Gromadzka *et al.*, 2016). The preferential deposition of ALYREF towards the 5' end of a transcript was also observed in iCLIP experiments (Viphakone *et al.*, 2019).

Furthermore, UAP56 and its paralog DDX39 interact with mRNA in an EJC and 5' cap-dependent manner (Gromadzka *et al.*, 2016). Of note, SR proteins can also link mRNAs with NXF1 and provide another connection between splicing and export (Muller-McNicoll *et al.*, 2016). Therefore, the EJC assists but is likely not essential for the export of spliced mRNAs.

1.2.4 Localization of mRNA

In highly polarized cells, the spatial enrichment of mRNAs is important to provide local protein synthesis. In order to be locally translated, exported mRNAs are actively transported as mRNPs by motor proteins along the cytoskeleton to their destination. This process has been intensely studied in the *Drosophila* oocyte (Weil, 2014). Furthermore, neuronal cells strongly depend on mRNA transport to establish a local proteome in dendrites and axons (Glock *et al.*, 2017, Das *et al.*, 2019).

The splicing-dependent deposition of all four EJC proteins has been shown to be essential for the correct localization of oskar mRNA in the posterior pole of the Drosophila embryo (Newmark & Boswell, 1994, Hachet & Ephrussi, 2001, Mohr *et al.*, 2001, van Eeden *et al.*, 2001, Palacios *et al.*, 2004). The transport occurs via kinesin I and microtubules (Pokrywka & Stephenson, 1995, Brendza *et al.*, 2000). How exactly the EJC is linked to mRNP transport is not yet clear. Besides the deposition of an EJC, the splicing of the first intron in the oskar mRNA leads to the formation of an RNA secondary structure which is also essential for transport (Ghosh *et al.*, 2012, Simon *et al.*, 2015). Whether an interplay between this structure and the neighboring EJC supports mRNP mobility remains to be investigated.

Although EJC components have been found in neuronal transport granules, there is no evidence so far that the EJC is essential for the localization of any other transcript than the oskar mRNA (Wang *et al.*, 2017).

1.2.5 Stimulation of translation and translation-dependent removal of the EJC

Spliced mRNAs yield more gene products in comparison to their intronless counterparts (Nott *et al.*, 2003). Part of this effect is due to splicing-dependent enhanced mRNA levels (Le Hir *et al.*, 2003). Furthermore, spliced mRNAs are translated more efficiently (Lu & Cullen, 2003). The EJC and peripheral interacting proteins have been shown to be largely responsible for the stimulatory effect of splicing on translation (Wiegand *et al.*, 2003, Nott *et al.*, 2004).

Several observations have been made that functionally link the EJC to translation. Firstly, it has been suggested that the EJC-interacting protein SKAR is involved in mediating the downstream translation stimulatory effect of mTOR signaling (Ma *et al.*, 2008). Secondly, it was reported that EIF4A3 is recruited to the 5' end of CBC-bound mRNAs and could potentially enhance translation by unwinding secondary structures in the 5' untranslated region (UTR) (Choe *et al.*, 2014). Thirdly, the EJC component CASC3 has been implicated to directly link the EJC to the translation machinery (Chazal *et al.*, 2013). CASC3 interacts with the translation initiation complex EIF3 and stimulates the translation of spliced mRNAs.

During the first round of translation, EJCs are normally removed by the processing ribosome (Dostie & Dreyfuss, 2002, Lejeune *et al.*, 2002). EJC disassembly might be aided by the ribosome-bound factor PYM (Diem *et al.*, 2007, Gehring *et al.*, 2009b). PYM has a strong affinity for the MAGOH-RBM8A heterodimer and presumably destabilizes mature EJCs via transient interactions with MAGOH-RBM8A (Gehring *et al.*, 2009b). Since the depletion of PYM results in a reduced translation efficiency of certain transcripts, it has been proposed that PYM also functions to promote the enhanced translation of EJC-bound transcripts (Diem *et al.*, 2007). Another interpretation is that PYM is crucial to recycle EJC core components and that it promotes translation by maintaining the available pool of EJC factors. This notion is supported by the fact that PYM is essential in *Drosophila* cells when levels of EJC components are reduced by a knockdown (Ghosh *et al.*, 2014). However, since PYM does not associate with the ribosome in *Drosophila*, the coupling of translation and disassembly by PYM might be limited to human cells (Ghosh *et al.*, 2014). In both *Drosophila* and human cells, the overexpression of PYM results in excess free PYM that destabilizes EJCs (Gehring *et al.*, 2009b, Ghosh *et al.*, 2014). Thus, the activity of PYM must be tightly regulated to prevent the global instability of EJCs.

After disassembly, the MAGOH-RBM8A heterodimer is reimported into the nucleus by importin-13 (Bono *et al.*, 2010). How EIF4A3 is transported back to the nucleus is not yet known.

1.2.6 EJC-dependent nonsense-mediated mRNA decay

During translation, the EJC fulfills an important role in mRNA surveillance. Transcripts that contain a premature termination codon (PTC) are degraded by the NMD pathway to prevent

the production of truncated and potentially toxic proteins. Since canonical termination codons are commonly located in the last exon, an EJC present more than 50 nucleotides downstream of a termination codon is a signal for NMD (Nagy & Maquat, 1998). Premature termination codons can arise for various reasons. Firstly, point mutations in the DNA or errors during transcription can turn coding codons into termination codons. Secondly, alternative splicing can lead to transcripts harboring a PTC. An example is the inclusion of intronic sequences with PTCs into the transcripts as poison cassette exons (Lareau *et al.*, 2007). Furthermore, frameshifts caused by alternative splicing can also lead to the introduction of PTCs. NMD is however not only a quality control pathway that prevents the accumulation of truncated proteins that would arise from erroneous transcript. Instead, by switching from a stable isoform to an NMD-sensitive variant via alternative splicing, the output of gene products can be regulated. This regulatory function is well described for the expression of SR proteins which are autoregulated by coupling alternative splicing and NMD (Lareau *et al.*, 2007, Ni *et al.*, 2007).

Many factors have been identified that are involved in EJC-dependent NMD (Figure 5). Via these proteins, the presence of a downstream EJC is communicated to a set of nucleases that degrade the targeted transcript. When the ribosome terminates at a PTC, the information of an EJC present downstream must be passed on to the NMD machinery. Any termination codon is recognized by the eukaryotic release factor 1 (eRF1) and eRF3 which trigger ribosome recycling and peptide release in a GTP-dependent manner (Schuller & Green, 2018). Based on interaction studies, a model was postulated by which the EJC is bridged to NMD factors via the SURF complex formed of SMG1-UPF1-eRF1-and eRF3 (Kashima *et al.*, 2006). According to this model, eRF1 and eRF3 at the translation termination codon recruit UPF1 which can be phosphorylated by SMG1. This phosphorylation is required for NMD (Yamashita *et al.*, 2005b). UPF3B directly interacts with the EJC with its exon junction complex binding motif (EBM) and bridges the EJC via UPF2 to UPF1. The phosphorylated UPF1 N- and C-termini recruit SMG5, SMG6 and SMG7 (Okada-Katsuhata *et al.*, 2012). The endonuclease SMG6 cleaves the mRNA in the vicinity of the PTC which results in a 5’ and 3’ fragment that are degraded involving the RNA exosome or the cytoplasmic 5’-3’ exonuclease XRN1 (Huntzinger *et al.*, 2008, Eberle *et al.*, 2009, Boehm *et al.*, 2014). The SMG5/7 heterodimer promotes the deadenylation of the transcript leading to decapping and 5’-3’ decay by XRN1 as well as 3’-5’ exonucleolytic degradation (Loh *et al.*, 2013). Analyses of the relative participation of these SMG6- and

SMG5/7-dependent pathways in the degradation of an NMD target have shown that they can act redundantly (Colombo *et al.*, 2017). Under normal conditions, SMG6-mediated endocleavage is the dominating pathway in human cells (Lykke-Andersen *et al.*, 2014, Schmidt *et al.*, 2015).

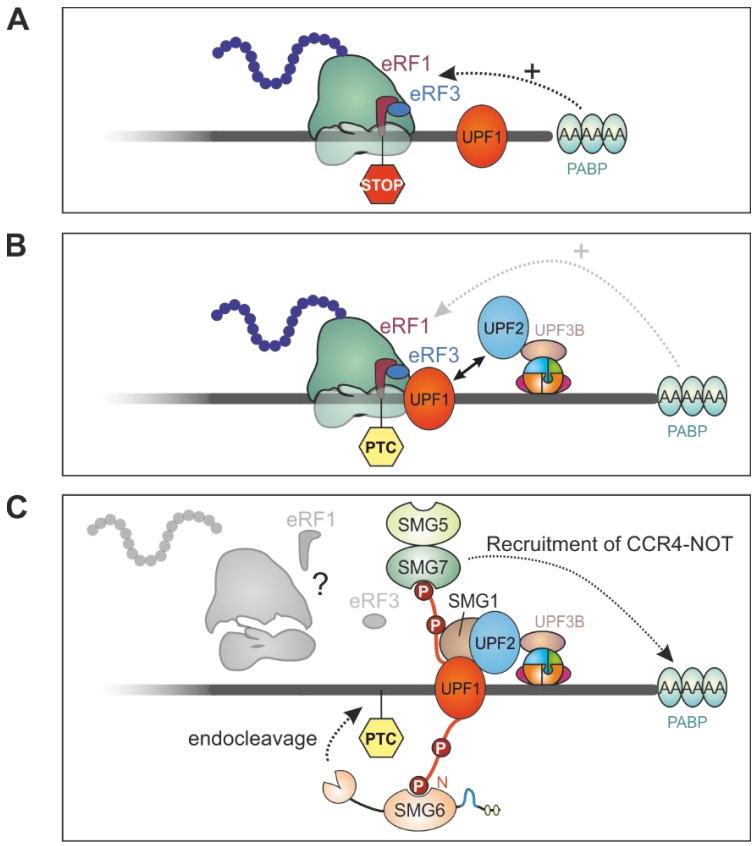


***Figure 5: Model of translation termination at a premature termination codon (PTC). A**: Translation termination at a canonical termination codon. The eukaryotic release factor 1 (eRF1) binds to the termination codon together with eRF3. The hydrolysis of GTP by eRF3 results in in its dissociation and subsequent steps that cause peptide release and ribosome recycling. The cytoplasmatic polyA-binding protein (PABP) stimulates translation termination. UPF1 is probably bound in the 3' UTR but remains inactive. **B**: Aberrant translation termination event leading to nonsense-mediated mRNA decay. A long distance between PABP and the termination codon reduces its stimulatory effect on termination. The stalling of the ribosome due to inefficient termination causes the activation of bound UPF1. An EJC present downstream strongly stimulates NMD potentially by acting as a binding platform for UPF3B which can be bridged to UPF1 by UPF2. **C**: The kinase SMG1 phosphorylates the N- and C-termini of UPF1. This phosphorylation can be caused by aberrant translation termination and/or by recruitment of UPF1. The phosphorylated UPF1 is bound by the endonuclease SMG6 which cleaves the mRNA at the termination codon. The UPF1 N-terminus is bound by the SMG5/SMG7 heterodimer which recruits CNOT8 (POP2), a component of the CCR4-NOT deadenlyase complex.*

Some observations cannot be explained by a strict SURF model relying on direct physical interactions. Firstly, it is not clear how long distances up to 600 nucleotides between the PTC and a downstream EJC can be spanned (Neu-Yilik *et al.*, 2001, Boehm *et al.*, 2014). Secondly,

the tethering of a UPF3B mutant which is unable to bind UPF2 is still able to induce NMD (Gehring *et al.*, 2003). This suggests that the NMD machinery can be recruited in an UPF2-independent manner. Finally, a combined knockdown of UP3B and its paralog UPF3A also does not inhibit NMD of the T-cell receptor-β mRNA, showing that an alternative route to communicate the presence of an EJC downstream of an EJC must exist (Chan *et al.*, 2007).

In tethering assays, more EJC interacting factors have been shown to be able to elicit NMD when bound downstream of a termination codon. Both RNPS1 and CASC3 strongly induce NMD when tethered to reporter mRNAs, but only RNPS1 requires UPF2 for this effect (Gehring *et al.*, 2005, Ciriello, 2014). This observation has led to the hypothesis that different branches of NMD activation exist. Whether these branches act on the same set of endogenous NMD targets or whether there are subsets of NMD targets with different requirements of factors is not clear. A recent study has claimed that RNPS1 is required for the degradation of many endogenous NMD targets, whereas CASC3 is not (Mabin *et al.*, 2018).

Many transcripts undergo NMD even though there is no EJC downstream of the termination codon (EJC-independent NMD). In transcripts with a long 3' UTR, there is large distance of the termination codon to the polyA binding protein (PABP). PABP has been shown to promote translation termination by promoting the GTP hydrolysis function of eRF3 and to counteract NMD when tethered closely downstream of a termination codon (Eberle *et al.*, 2008). Furthermore, in an *in vitro* translation termination system, PABP stimulates translation termination by enhancing the recruitment of eRF1 and eRF3 to the ribosome (Ivanov *et al.*, 2016). Interestingly, in transcripts with a long 3' UTR endocleavage also occurs at the PTC and is strictly UPF2-dependent (Boehm *et al.*, 2014). How the NMD machinery is recruited to transcripts with long 3' UTRs is not known. The 3' UTR length was shown not to be a predictor of NMD sensitivity per se since many transcripts with this feature are insensitive to NMD (Singh *et al.*, 2008). A subset of these contain AU-rich sequences directly downstream of the termination codon and which render them immune to NMD (Toma *et al.*, 2015).

Although UPF1 is regarded as the central NMD factor, UPF1-binding in itself is not a trigger for NMD since it was found to bind promiscuously to the coding sequence and 3' UTR of mRNAs (Zund *et al.*, 2013). A common denominator in all models of NMD is rather that a translation termination event is recognized as aberrant which leads to the activation of the NMD machinery. In studies performed in yeast and human cells, it was shown that the ribosome

remains longer on PTCs that trigger NMD (Amrani *et al.*, 2004, Peixeiro *et al.*, 2012). This delay could potentially signal an aberrant termination event and give time for the recruitment of UPF1 by the release factors and subsequent phosphorylation. A downstream bound EJC and the distance to PABP could either enhance or decrease this process. UPF3B has also been shown to interact *in vitro* with the release factors and negatively impact translation termination, thus establishing a link between the EJC and the termination event (Neu-Yilik *et al.*, 2017).

For a long time, the prevalent view was that NMD would occur exclusively during the first round of translation (Schoenberg & Maquat, 2012). Directly after export, transcripts undergo the initial round of translation while bound by the CBC (Ryu & Kim, 2017). Either during translation or translation-independently, the CBC is replaced by eIF4E which forms the translation initiation complex eIF4F together with eIF4A and eIF4G (Sato & Maquat, 2009). There is evidence however that EIF4E-bound transcripts are also NMD competent (Durand & Lykke-Andersen, 2013, Rufener & Muhlemann, 2013). Supporting this notion, a recent study based on live-cell imaging of transcripts degraded by NMD has reported that transcripts undergo on average eight rounds of translation before endocleavage occurs (Hoek *et al.*, 2019). This could be explained by a model in which NMD is a stochastic process and the probability of its activation increases or decreases depending on the mRNP composition downstream of the termination codon. This recent study also confirmed an earlier observation that a subset of transcripts within an NMD-targeted mRNA pool can escape NMD (Trcek *et al.*, 2013, Hoek *et al.*, 2019). Thus, a variability of NMD efficiency seems to exist between subpopulations of mRNA.

1.3 Monitoring the dynamics of mRNA turnover with viral sequence elements

The amount of a given transcript in the cell is determined by its rate of transcription and degradation. The turnover of mRNAs is therefore a crucial process to enable the regulation of gene expression. It also provides a way by which faulty transcripts or exogenous infectious RNAs can be removed. The stability of mRNAs can be controlled by many features such as the length of its polyA tail, its 5’ cap, sequence or secondary structure elements recognized by small RNAs or RNA-binding proteins, and quality control mechanisms. mRNA decay involves many nucleolytic pathways acting on a transcript and is often linked to other cellular processes

such as transcription or translation. For these reasons, mRNA decay measurements that do not disturb cellular processes are helpful to detangle the relative contribution of these pathways on the degradation of a given transcript. This chapter describes methods to study mRNA degradation and how viral exonuclease-resistant RNA structures could improve their readout.

1.3.1 mRNA turnover

Both in the nucleus and in the cytoplasm, distinct mRNA decay machineries exist that act in canonical mRNA degradation and mRNA quality control mechanisms. The nuclear RNA decay machinery consists of the nuclear 5'-3' exonuclease XRN2 and the RNA exosome which possesses 3'-5' exonuclease and endonuclease activity (Schmid & Jensen, 2018). Nuclear RNAs that are targeted towards degradation can arise for many reasons. As described above, XRN2 functions during transcription in degrading the cleavage fragment caused by 3' processing. Furthermore, it contributes to 5' trimming of ribosomal pre-RNAs (Wang & Pestov, 2011). Pervasive transcription and mistakes during RNA processing lead to the formation of many unstable RNA species that will be degraded by the RNA exosome (Peck *et al.*, 2019).

Bulk degradation of mature mRNAs in the cytoplasm starts by shortening and removing the polyA tail (Chen & Shyu, 2011). Two major deadenylation complexes shorten polyA tails in mammalian cells: PAN2-PAN3 and CCR4-NOT (Wolf & Passmore, 2014, Collart, 2016). A biphasic model according to which PAN2-PAN3 targets longer polyA tails and CCR4-NOT subsequently takes over has been assumed for a long time (Yamashita *et al.*, 2005a). Recent results suggest that the bulk of deadenylation is carried out by CCR4-NOT and that the catalytic subunit CCR4 acts on PABP-bound polyA segments whereas the other catalytic subunit CAF1 targets free polyA stretches (Yi *et al.*, 2018). Results obtained in yeast have supported this cyclic model of deadenylation (Webster *et al.*, 2018).

Deadenylation promotes the removal of the 5' cap by the decapping complex. DCP2 removes the cap and is aided by its activator DCP1 and the scaffold protein EDC4 (Chang *et al.*, 2014a). The exposed 5' monophosphate end is then vulnerable to exonucleolytic cleavage by the cytoplasmic 5'-3' exonuclease XRN1 (Nagarajan *et al.*, 2013). Deadenylated transcripts are also a target for degradation via the RNA exosome (Mitchell *et al.*, 1997, Wang & Kiledjian, 2001). Transcripts with a short polyA tail can be uridylated by the terminal uridylyl transferases

(TUTases) TUT4 and TUT7 (Chang *et al.*, 2014b, Lim *et al.*, 2014). PABP prevents TUT7 from uridylating polyA tails longer than 25 nts (Lim *et al.*, 2014). Uridylated mRNAs undergo decay from the 5′ end by decapping and XRN1-dependent decay and from the 3′ end by the exosome and the uridylation-specific 3′-5′ exonuclease DIS3L2 (Lim *et al.*, 2014, Thomas *et al.*, 2015).

Sequences or secondary structures that influence the stability of an mRNA are commonly located in the 3′ UTR (Bartel, 2009, Uchida *et al.*, 2019). Small RNAs such as miRNAs, siRNAs or piwi RNAs (piRNA) are complementary to target sequences in transcripts (Ha & Kim, 2014). Together with members of the AGO protein family, they form RNA-induced silencing complexes (RISC). The binding of the RISC to transcripts promotes their translational repression or degradation (Huntzinger & Izaurralde, 2011). While miRNAs and siRNAs commonly regulate protein levels, piRNAs specifically function in germ cells (Ha & Kim, 2014).

Another type of sequence element that influences transcript stability is the AU-rich element. AU-rich elements are recognized by RNA-binding proteins that will either increase or decrease the stability of the mRNA (Garcia-Maurino *et al.*, 2017). AU-rich elements can cause rapid degradation via the TIS11 family of RNA-binding proteins, namely TTP and BRF1 and 2 (Sanduja *et al.*, 2011). The mode of action is swift deadenylation followed by the degradation of the transcript (Wilson & Treisman, 1988). AU-rich elements of this type are a feature of many transcripts involved in cell proliferation or immunity which need to be tightly regulated (Anderson, 2008). In contrast, the human antigen R (HuR) protein competes with degradation promoting factors and generally has an enhancing effect of mRNA stability (Fan & Steitz, 1998). RBPs that recognize AU-rich elements can also regulate each other and their own transcripts, thus forming a complex network (Garcia-Maurino *et al.*, 2017).

Besides sequence elements, secondary structures can influence transcript stability. This type of structure was discovered in the transcripts encoding the ferritin and transferritin receptor proteins which function in iron uptake and storage (Hentze *et al.*, 1987, Mullner & Kuhn, 1988). Five Iron-response elements formed of 30 nt long hairpins in the hTR 3′ UTR lead to the degradation of hTR mRNA by endocleavage (Pantopoulos, 2004). The identity of the endonuclease is not known (Schoenberg, 2011). When iron levels are low, the hTR transcript is stabilized by iron response proteins (IRPs) binding to the IREs, thus promoting iron uptake (Mullner *et al.*, 1989). The expression of ferritin which stores iron in the cell is also regulated

by an IRE. An IRE in the 5' UTR of ferritin mRNA is bound by IRP when iron levels are low and prevents translation initiation (Gray & Hentze, 1994).

Another type of stem loop that promotes mRNA degradation is the constitutive decay element (CDE) found in the 3' UTR of cytokine transcripts such as TNF-α and interleukin-6 (Mino *et al.*, 2015). This structure is recognized by Roquin1 and 2 which recruit the CCR4-NOT deadenylation complex (Leppek *et al.*, 2013). Furthermore, the endonuclease ZC3H12A cleaves the transcript in the vicinity of the stem loop (Mino *et al.*, 2015).

Codon optimality is another feature of mRNA that influences its stability. Because certain amino acids can be encoded using distinct tRNAs, the restricted availability of these charged tRNAs potentially slows down translation. A recent study showed that transcripts with a non-optimal codon composition are less stable (Wu *et al.*, 2019b).

Transcript stability is not only controlled by inherent sequences recognized by RNA-binding proteins, but also quality control processes coupled to translation. This includes the aforementioned nonsense-mediated mRNA decay mechanism that targets transcripts with a premature termination codon. Studied mainly in yeast, the no-go decay (NGD) pathway occurs when ribosomes encounter blocks such as hairpins or A-stretches during translation which lead to ribosome stalling and colliding (Simms *et al.*, 2017). Similarly to NMD, converging nucleolytic pathways act on the decay substrate. The endonuclease Cue2 has been identified in yeast to cleave transcripts at the position of colliding ribosomes (D'Orazio *et al.*, 2019). Exonucleolytic decay especially by XRN1 is however the dominant contributor to the decay of NGD targets. Transcripts that completely lack a termination codon are thought to be degraded by the non-stop decay (NSD) pathway. Because the translation of the polyA tail into poly-lysines causes ribosome stalling, NSD might resemble NGD (Shoemaker & Green, 2012). It is not known whether NGD or NSD have functions like NMD that go beyond quality control in the regulation of gene expression. Since many transcripts are alternatively polyadenylated it is possible that some of these will be targeted by NGD (Shoemaker & Green, 2012).

The modification of RNA bases beyond the 7-methylguanosine cap have also been reported to influence transcript stability. N^6-methyladenosine (m^6A) is the most abundant internal RNA modification and has been linked to mRNA instability (Sommer *et al.*, 1978, Ke *et al.*, 2017). This reversible modification is added by the m^6A methyltransferase METTL3 and removed by the m^6A demethylase ALKBH (Bokar *et al.*, 1997, Zheng *et al.*, 2013). The YTHFD2 protein

serves as a reader for m^6A and has been proposed to induce decay by binding to m^6A-methylated mRNA and promoting its location to P-bodies (Wang *et al.*, 2014a). P-bodies are membrane-less subcompartments in the cytoplasm which contain mRNA and mRNA decay factors and have been implicated as sites of mRNA decay (Luo *et al.*, 2018). This view has been challenged however by the observation that preventing the assembly of P-bodies does not lead to a reduction in mRNA decay (Eulalio *et al.*, 2007). m^6A modifications recognized by YTHFD1 have been reported to enhance translation efficiency (Wang *et al.*, 2015). Another link to translation comes from the observation that m^6A modifications in the 5′ UTR enable cap-independent translation (Meyer *et al.*, 2015). It has been proposed that like for AU-rich elements, m^6A modifications can result in different outcomes depending which factors read them (Wang *et al.*, 2015).

Another RNA modification linked to mRNA stability is found on transcripts in which the first nucleotide adjacent to the cap is an adenosine. As in all mRNAs, the first nucleotide is methylated at 2′-hydroxide position as part of the cap structure (Ramanathan *et al.*, 2016). In up to 75% of mRNAs this A_m is additionally modified at the N^6 position, resulting in a N^6,2′-O-dimethyladenosine (m^6A_m) (Mauer *et al.*, 2017, Akichika *et al.*, 2019). This modification impairs DCP2's ability to remove the 5′ cap, thus potentially increasing the stability of a transcript (Mauer *et al.*, 2017). Two different studies analyzing a knockout of the responsible methyltransferase PCIF1 have come to divergent conclusions with only one of them reporting an upregulation of a subset of m^6Am-containing transcripts upon PCIF1 depletion (Akichika *et al.*, 2019, Boulias *et al.*, 2019). The demethylase FTO has been shown to remove m^6A_m modifications, suggesting it could be a modifier of mRNA stability. Recent evidence however has established that the predominant role of FTO lies in the modification of small nuclear RNAs involved in splicing (Mauer *et al.*, 2019).

1.3.2 Viral XRN1-resistant RNAs

Viruses of the genus *flavivirus* have developed a specialized secondary structure that protects a part of their genome from the degradation of the host's cytoplasmic XRN1 (Figure 6) (Pijlman *et al.*, 2008, Funk *et al.*, 2010, Silva *et al.*, 2010). These viruses possess a single-stranded positive sense RNA genome consisting of a 5′ UTR, an open reading frame coding for viral structure and replication proteins as well as a 3′ UTR (Chong *et al.*, 2019). At the beginning of the 3′ UTR they carry one or two conserved stem loops that fold into a pseudoknot structure.

Further downstream, one or more dumbbell-like secondary structures enhance the inhibition of XRN1. These structures make the 3' UTR resistant to degradation by the host's cellular XRN1. The stabilized 3' UTR can then be processed into functional non-coding RNAs. These small RNAs, termed subgenomic flaviviral RNAs (sfRNAs), have been shown to inhibit the host's RNAi response and type-I-interferon response, thereby harming anti-viral defence mechinsms (Schnettler *et al.*, 2012, Schuessler *et al.*, 2012). Additionally, sfRNAs bind to RNA-binding proteins involved in post-transcriptional processes such as splicing, RNA decay and translation (Michalski *et al.*, 2019). Also, because XRN1 is stalled at the 3' UTR, these sequences potentially trap XRN1 there, thus reducing the host's ability to degrade infectious RNAs (Michalski *et al.*, 2019).

Flaviviruses are unique in the fact that they hijack the host's mRNA decay machinery to produce pathogenic RNAs (Slonchak & Khromykh, 2018). The genus contains species such as Zika, West Nile, Dengue and Yellow fever virus which are highly pathogenic for humans and cause outbreaks affecting millions (Chong *et al.*, 2019). The investigation of features that confer their pathogenicity has therefore attracted great interest in the recent years.

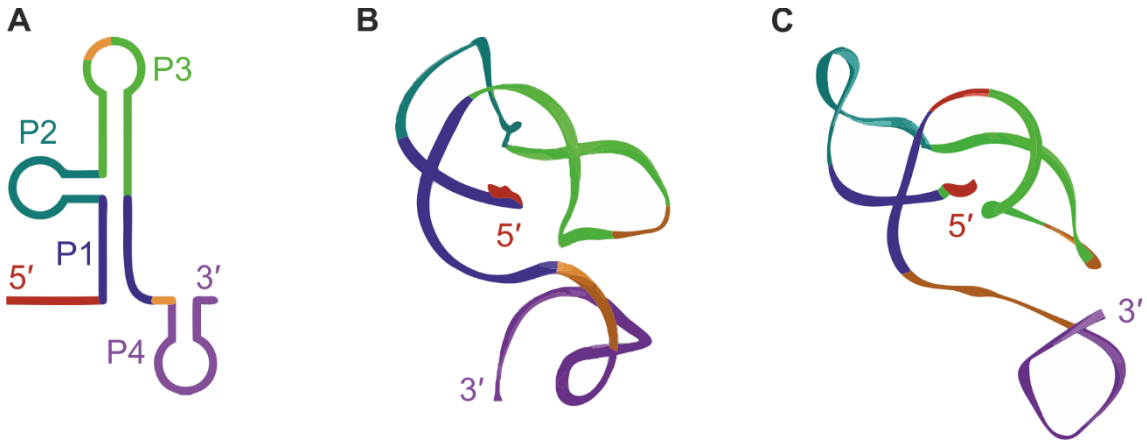


***Figure 6: Structures of XRN1-resistant RNAs. A**: Schematic depiction of one XRN1-resistant element formed of three stem-loops. **B**: Structure of the Murray Valley Encephalitis (MVE) xrRNA2 (pdb: 4PQV). The 5' end of the RNA passes through a ring which halts the progression of XRN1 **C**: Structure of the Zika virus xrRNA1 (pdb: 5TPY). In comparison to the MVE structure the knot is in a tighter conformation. The structures were rendered with Chimera v.1.14rc.*

In 2014 the structure of the Murray Valley Encephalitis (MVE) virus's xrRNA was solved, laying the foundations of understanding the mechanism of XRN1-inhibition (Figure 6B) (Chapman *et al.*, 2014). The 5' end of the xrRNA is surrounded by a ring-like structure which is stabilized by a three-way junction as well as interactions of the 5' end with the ring itself. Furthermore, interactions between P3 and P4 (Figure 6, depicted in orange) could further stabilize this

pseudoknot fold. However, in the MVE structure they are not base-paring. The ring formed of P1 and P3 physically blocks XRN1-progression from the 5' end (Kieft *et al.*, 2015). The mechanism of inhibition is specific to the 5' end, since the viral RNA-polymerase is able to progress in 3' to 5' direction during RNA replication (Slonchak & Khromykh, 2018). The xrRNA structure obtained from the Zika virus confirmed the pseudoknot structure (Figure 6C) (Akiyama *et al.*, 2016). Here, interactions between P3 and P4 are stabilized by stacking of the P4 helix, thus forming a fully folded xrRNA.

1.3.3 RNA decay measurements

Several challenges are linked to measuring mRNA degradation. The degradation of an mRNA has to be measured independently of its rate of synthesis. Furthermore, an essential challenge in studying mRNA decay is that the transcript escapes detection when it has been degraded. This makes it necessary to measure the half-lives transcripts or to use techniques that enable the detection of decay intermediates (Wada & Becskei, 2017). Finally, mRNA decay measurements often require interfering with endogenous cellular processes which could potentially influence the readout of the assay.

The degradation of a transcript can be measured by studying the amount of mRNAs of the same type. A pool of mRNAs can be directly detected on a northern blot using a labelled probe. This technique visualizes all RNA species that carry the probe binding site including pre-mRNA or alternatively spliced variants. The design of specific probes and the use of radioactive labelling can be an obstacle when using this method. More common is the detection by quantitative real-time PCR (qPCR) which involves reverse transcribing the RNA into cDNA before measuring. This method is well-established in most molecular biology labs. Drawbacks include that the primers used have to be vigorously tested before allowing quantitative readouts (Bustin *et al.*, 2009).

When studying an endogenous mRNA species, the amount is determined both by the rate of synthesis and decay. Transcription can change dynamically due to the regulation by transcription factors. To specifically study the decay kinetics of degradation, several options to study half-life times of mRNAs exist. Newly produced mRNAs can be labelled, either using radioactive or non-radioactive nucleotides. Transcriptome-wide measurements of mRNA decay are also possible: e.g. the SLAM-seq method uses a pulse of 4-thiuridine labeling to

distinguish between newly synthesized mRNAs and calculate transcript half-lives (Herzog *et al.*, 2017). Both radioactive and non-radioactive types of nucleotides can negatively impact the cells metabolism and therefore potentially impact mRNA decay measurements: Phosphorous 32 can cause DNA damage and 4-thiouridine which is mostly used in mammalian cells disturbs rRNA processing thus leading to nucleolar stress (Burger *et al.*, 2013, Cheng *et al.*, 2015). Another approach is to block the synthesis of new transcripts by treating cells with transcription inhibitors (Wada & Becskei, 2017). By quantifying the change of existing mRNA molecules their half-lives can be determined. A wide array of transcriptional inhibitors with different selectivity for RNA polymerases and speed of action exist (Bensaude, 2011). Because global transcription is blocked, this treatment can have generally adverse effects on cells (Wada & Becskei, 2017).

Reporter mRNAs that underlie the transcriptional control of an adjustable promoter are a useful tool to study mRNA turnover. These constructs can either be stably integrated into a genomic locus by stable transfection or transiently expressed from extra-chromosomal expression plasmids. Two types of promoters exist for these kinds of studies: firstly the transcription can be turned on, e.g. in Tet-On systems (Gossen & Bujard, 1992, Gossen *et al.*, 1995). Because the rate of transcription is estimated to be constant under these conditions, steady state levels of the transcript can be used as a read-out. Promoters that are turned off by the addition of a substance, e.g. Tet-Off systems can be used to halt the expression of a reporter mRNA so that the half-life of remaining transcript levels can be determined (Rennel & Gerwins, 2002).

Finally, it is possible to visualize mRNA degradation by using imaging approaches. Single-molecule fluorescent in situ hybridization (smFISH) has revealed insights into mRNA decay mechanisms of human, yeast and trypanosome cells (Trcek *et al.*, 2011, Trcek *et al.*, 2013, Kramer, 2017). Imaging the turnover of mRNA is also possible in real-time by using arrays of binding sites within the mRNA that are bound by fluorescent proteins (Bertrand *et al.*, 1998). This has recently been combined with the SunTag system which provides a visual readout of translation (Tanenbaum *et al.*, 2014, Yan *et al.*, 2016, Hoek *et al.*, 2019).

Another common technique of investigating mRNA turnover is a knockdown of factors involved in mRNA degradation and measuring the stabilization of their targets. A knockdown of XRN1 for example stabilizes decay intermediates resulting from endocleavage, making it

possible to map the cleavage sites (Lykke-Andersen *et al.*, 2014). Many pathways that target unstable mRNAs are linked however, resulting in the perturbation of not only a single element. For example, XRN1 directly interacts with EDC4 of the decapping complex, the decapping activator PatL1 and with the CCR4-NOT deadenylation complex (Braun *et al.*, 2012, Chang *et al.*, 2019).

1.4 Aims of this work

1. Although more than 200 ribosomal biogenesis factors in the nucleolus have been described in human cells, it is possible that biogenesis factors located outside of the nucleolus have not yet been identified. In embryonic stem cells UBR5 is upregulated and is functionally linked to maintain pluripotency. In Saez *et al.* (2020) an interaction between UBR5 and the H/ACA snoRNP in stem cells is reported. This has potential implications for rRNAs as the targets of H/ACA pseudouridylation. In this study the consequences of UBR5 depletion on rRNA maturation and ribosomal biogenesis are investigated.

2. During splicing, the spliceosome undergoes dramatic compositional changes. EJC deposition is known to be splicing-dependent, but structural and temporal details about this process have been unclear. The recently published cryo-EM structures of spliceosome intermediates have resolved the position of the EJC in relation to the spliceosome for the first time. In Gerbracht & Gehring (2018) these structures are described with a focus on newly identified interactions between EJC components and spliceosomal proteins. In Boehm *et al.* (2018) the mechanism by which the deposition of an EJC influences the splicing of neighboring junctions is elucidated. Although CASC3 is generally described as an essential core component of the EJC in the literature, we did not see any mis-splicing when we depleted CASC3 in human cells. In Gerbracht *et al.* (2019) CASC3 knockout cell lines are used to address the reported functions of CASC3 in splicing and the nonsense-mediated mRNA decay mechanism.

3. Often, multiple nucleolytic pathways are involved in the degradation of a transcript. Inserting viral XRN1-resistant RNAs into reporter mRNAs should selectively block 5’-3’ degradation and allow for the detection of decay intermediates at the same time. Since the accumulation of XRN1-resistant degradation intermediates directly corresponds to 5’-3’ degradation, they represent a readout of the contribution of 5’-3’ decay. In Boehm *et al.* (2016) multiple decay pathways are analyzed using xrRNAs as a proof of concept. Gerbracht *et al.* (2017) employs xrRNA-based analyses to study the effects of transfection methods on the readout of nonsense-mediated decay reporter mRNAs. Voigt *et al.* (2019) provides a step-by-step protocol to study mRNA degradation using xrRNA both with biochemical assays and single-molecule imaging tools.

2 PUBLICATIONS

- Saez, I, Gerbracht, JV, Koyuncu, S, Lee, HJ, Horn, M, Kroef, V, Denzel, MS, Dieterich, C, Gehring, NH and Vilchez, D (2020) **The E3 ubiquitin ligase UBR5 interacts with the H/ACA ribonucleoprotein complex and regulates ribosomal RNA biogenesis in embryonic stem cells.** FEBS Lett, 594, 175-188.

- Gerbracht JV, Gehring NH (2018) **The exon junction complex: structural insights into a faithful companion of mammalian mRNPs.** Biochem Soc Trans 46: 153-161

- Boehm V, Britto-Borges T, Steckelberg AL, Singh KK, Gerbracht JV, Gueney E, Blazquez L, Altmuller J, Dieterich C, Gehring NH (2018) **Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity.** Mol Cell 72: 482-495 e7

- Gerbracht JV, Boehm V, Britto-Borges T, Kallabis S, Wiederstein JL, Ciriello S, Aschemeier DU, Krüger M, Frese CK, Altmüller J, Dieterich C, Gehring NH (2019) **CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex.** bioRxiv: 811018

- Boehm V*, Gerbracht JV*, Marx MC, Gehring NH (2016) **Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences.** Nat Commun 7: 136

- Gerbracht JV, Boehm V, Gehring NH (2017) **Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells.** Sci Rep 7: 10616

- Voigt F*, Gerbracht JV*, Boehm V, Horvathova I, Eglinger J, Chao JA, Gehring NH (2019) **Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts.** Nat Protoc 14: 1603-1633

* These authors contributed equally

FEBS Letters

FEBS PRESS science publishing by scientists

The E3 ubiquitin ligase UBR5 interacts with the H/ACA ribonucleoprotein complex and regulates ribosomal RNA biogenesis in embryonic stem cells

Isabel Saez[1], Jennifer V. Gerbracht[2], Seda Koyuncu[1], Hyun Ju Lee[1], Moritz Horn[3], Virginia Kroef[3], Martin S. Denzel[3], Christoph Dieterich[4], Niels H. Gehring[2] and David Vilchez[1]

1 Institute for Genetics and Cologne Excellence Cluster for Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Germany
2 Institute for Genetics, Department of Biology, University of Cologne, Germany
3 Max Planck Institute for Biology of Ageing, Cologne, Germany
4 Section of Bioinformatics and Systems Cardiology, Department of Internal Medicine III and Klaus, Tschira Institute for Computational Cardiology, University Hospital, Heidelberg, Germany

Correspondence
D. Vilchez, Cologne Excellence Cluster for Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Joseph Stelzmann Strasse 26, 50931 Cologne, Germany
Tel: +49 22147884172
E-mail: dvilchez@uni-koeln.de

(Received 4 June 2019, revised 24 July 2019, accepted 26 July 2019, available online 8 August 2019)

doi:10.1002/1873-3468.13559

Edited by Michael Ibba

UBR5 is an E3 ubiquitin ligase involved in distinct processes such as transcriptional regulation and development. UBR5 is highly upregulated in embryonic stem cells (ESCs), whereas its expression decreases with differentiation, suggesting a role for UBR5 in ESC function. However, little is known about how UBR5 regulates ESC identity. Here, we define the protein interactome of UBR5 in ESCs and find interactions with distinct components of the H/ACA ribonucleoprotein complex, which is required for proper maturation of ribosomal RNA (rRNA). Notably, loss of UBR5 induces an abnormal accumulation of rRNA processing intermediates, resulting in diminished ribosomal levels. Consequently, lack of UBR5 triggers an increase in p53 levels and a concomitant decrease in cellular proliferation rates. Thus, our results indicate a link between UBR5 and rRNA maturation.

Keywords: embryonic stem cells; protein–protein interactions; ribosomal RNA; ribosome; ubiquitin ligases

UBR5 is an E3 ubiquitin ligase which belongs to the HECT (homologous to the E6-AP carboxy terminus) ligase family [1]. This 300 kDa-enzyme is mostly localized in the nucleus [2] and contains several functional domains. Besides its HECT-ubiquitin ligase domain, UBR5 also has a UBR box finger-like domain, which targets proteins containing the destabilizing N-degron signal toward degradation [3]. Moreover, it contains a PABC [polyadenylate-binding protein (PABP) C terminus] domain, also known as MLLE domain, which confers UBR5 the ability to bind mRNA and modulate eukaryotic translation initiation [4]. Notably, UBR5 exhibits E3-independent activity as a transcriptional cofactor for the progesterone receptor and acts as a binding partner for distinct proteins such as GW182, CHK2, and DUBA [5–7]. UBR5 was first discovered as a tumor suppressor in a progestin-modifying screening using breast cancer cells [1]. In these lines, several studies evidenced a role of UBR5 in cell cycle progression and tumorigenesis [8,9]. UBR5 is also an important mediator of the DNA damage response, since it is a substrate of DNA damage response kinases such as ATM [10,11]. Furthermore, UBR5 has a role in transcriptional regulation, as it

Abbreviations
GO, gene ontology; hESCs, human embryonic stem cells; mESCs, mouse embryonic stem cells; PABP, polyadenylate-binding protein; RIP, RNA immunoprecipitation; RNP, ribonucleoprotein; rRNA, ribosomal RNA; snoRNAs, small nucleolar RNAs; snRNA, small nuclear RNA.

binds and promotes the transactivation of the progesterone receptor [2], ubiquitinates the CDK9 subunit of the positive transcription elongation factor b [12], and regulates the miRNA pathway [13]. In addition, *Ubr5*$^{-/-}$ mice embryos are not viable, indicating that this protein is essential for development [14].

Recently, we have shown that UBR5 is highly expressed in human embryonic stem cells (hESCs) and its expression decreases when hESC differentiate, suggesting a role of UBR5 in ESC identity [15]. In these lines, a study reported that *Ubr5* knockdown results in significant loss of pluripotency markers in mouse ESCs (mESCs) [16]. Here we examine the role of UBR5 in ESCs by defining its interactome using immunoprecipitation assays followed by quantitative proteomics. We find that UBR5 interacts with the H/ACA ribonucleoprotein (RNP) complex, which participates in the proper maturation and processing of ribosomal RNA (rRNA). We further characterize the role of UBR5 in this system by analyzing pre-rRNA processing and find that the levels of several pre-rRNA intermediates are altered in *Ubr5*$^{-/-}$ ESCs. Finally, we show a decreased translational profile in UBR5-defective cells accompanied by increased p53 levels and diminished proliferation rates.

Methods

hESC and mESC lines and culture

The H9 (WA09) hESC line was obtained from the WiCell Research Institute and was maintained on Geltrex (ThermoFisher Scientific, Waltham, MA, USA) using mTeSR1 (Stem Cell Technologies, Cologne, Germany). Undifferentiated hESC colonies were passaged using a solution of dispase (2 mg·mL^{-1}), and scraping the colonies with a glass pipette. The mESC AN-13 mESC line was obtained from the Haplobank at the Institute of Molecular Biotechnology GmbH. The mESC was cultured in noncoated plates using DMEM supplemented with 7.25% FBS, 1% Pen/Strep, 1% nonessential amino acids, 1% Glutamax, 1% Sodium pyruvate, 50 μM β-mercaptoethanol and human LIF (12 ng/mL). The cell lines used in this study were tested for mycoplasma contamination at least once every three weeks. No mycoplasma contamination was detected. Research involving hESCs was performed with approval of the German Federal competent authority (Robert Koch Institute).

Lentiviral infection of hESCs

Lentivirus (LV)-nontargeting shRNA control, LV-UBR5 shRNA #1 (TRCN0000003411), and LV-UBR5 shRNA #2 (TRCN0000226458) in pLKO.1-puro vector were obtained from Mission shRNA (Sigma, Hamburg, Germany). Transient infection experiments were performed as follows. H9 colonies growing on Geltrex were individualized using Accutase. Hundred thousand cells were plated on Geltrex plates and incubated with mTesR1 medium containing 10 μM ROCK inhibitor for 1 day. Then, cells were infected with 5 μL of concentrated lentivirus. Cells were fed with fresh media the day after to remove the virus. After 1 day, cells were selected for lentiviral integration using 2 μg·mL^{-1} puromycin (ThermoFisher Scientific).

CRISPR/Cas9- mediated KO of mESCs

UBR5 gene knockout was done in mouse embryonic stem cells carrying a haploid chromosome set [[17]]. These cells become diploid over time and mutations thus remain homozygous. For genome engineering CRISPR/Cas9 was used as previously described [18]. Sequences for small guide RNAs were designed online (http://crispor.org) and purchased from Sigma ($F^{Guide\ 1}$: CACCGTAAATGATTTAC CATACGGT, $R^{Guide\ 1}$: AAACACCGTATGGTAAATCA TTTAC, $F^{Guide\ 2}$: CACCGATTGCTTTAAAACTCCAC TT, $R^{Guide\ 2}$: AAACAAGTGGAGTTTTAAAGCAATC). The primers were cloned into the Cas9-GFP expressing plasmid PX458 (Addgene #48138, gift from Feng Zhang). A combination of guides and the Cas9-expressing plasmid were transfected using lipofectamine 3000 (Thermo Fisher Scientific) according to manufacturer's instructions. Twenty-four hours post transfection a mixed population of haploid and diploid ES cells were stained with 10 μg/mL Hoechst 33342 (Thermo Fisher Scientific) for 30 min and sorted according to their DNA content and GFP expression on a FACSAria Fusion sorter (BD, Franklin Lakes, NJ, USA). Haploid GFP-positive cells were single cell sorted into 96-well plates. Emerging clones were transferred to 24-well plates 7 days later and genotyped following DNA extraction (DNA extraction solution, Epicentre Biotechnologies, Madison, WI, USA) using the following primers; F: GAGAC CCGCCTGTTTGTTTT, R: CCCAATTGATTCTCTGAG CCA. Sanger sequencing of PCR products was performed at Eurofins Genomics GmbH, Ebersberg, Germany. UBR5 knockout clones were selected for diploid cells (FACS after Hoechst staining) prior to further experiments.

Transfection of HEK293T cells

HEK293T cells (ATCC) were plated on 0.1% gelatin-coated plates and grown in DMEM supplemented with 10% FBS and 1% MEM nonessential amino acids (ThermoFisher Scientific). Cells were transfected once they reached 80-90% confluency. About 1 μg GFP-UBR5 wild-type or GFP-UBR5 ΔHECT overexpression plasmid and 1 μg of GAR1-FLAG, NHP2-FLAG or DKC1-FLAG were used for transfection, using Fugene HD

(Promega, Mannheim, Germany) following manufacturer's instructions. After 24 h incubation in normal medium, the cells were harvested for further experiments. GFP-UBR5 wild-type and GFP-UBR5 ΔHECT overexpression plasmids (Addgene plasmids #52050 and # 52051, respectively) were a gift from D. Saunders and were first published in ref [19]. DKC1, GAR1, and NHP1 were PCR-amplified from HeLa cDNA and cloned between *Xho I*/*Not I* into a modified pCI-neo (Clontech, Saint-Germain-en-Laye, France) plasmid, which contains an N-terminal FLAG-tag sequence (pCI-FLAG).

Immunoprecipitation of UBR5 for interactome analysis

The immunoprecipitation was performed under the conditions described in [20]. Briefly, hESCs were lysed in RIPA buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% Triton x-100, 1% sodium deoxycholate, 1 mM EDTA, 1 mM PMSF) supplemented with protease inhibitor cocktail (Roche) and centrifuged at 13 000 ***g*** for 15 minutes at 4 °C. About 350 μg of protein was incubated with UBR5 Antibody (Cell Signaling, Frankfurt am Main, Germany, #8755, 1 : 50) or FLAG antibody as a control (Sigma, F7425, 4 μg). Subsequently, samples were incubated with 100 μL of μMACS Micro Beads for 1 h at 4 °C and loaded to precleared μMACS column (#130-042-701). Beads were washed three times with 50 mM Tris (pH 7.5) buffer containing 150 mM NaCl, 5% glycerol, and 0.05% Triton and then washed five times with 50 mM Tris (pH 7.5) and 150 mM NaCl. Then, columns were subjected to in-column tryptic digestion containing 7.5 mM ammonium bicarbonate, 2 M urea, 1 mM DTT, and 5 $ng{\cdot}mL^{-1}$ trypsin. Digested peptides were eluted using 50 μL of elution buffer 1 containing 2 M urea, 7.5 mM Ambic, and 5 mM IAA two times. Digests were incubated over night at room temperature and samples were stage-tipped the next day for label-free quantitative proteomics. All samples were analyzed on a Q-Exactive Plus (Thermo Scientific) mass spectrometer that was coupled to an EASY nLC 1200 UPLC (Thermo Scientific). All mass spectrometric raw data were processed with Maxquant and the resulting output was processed using Perseus. Significant differences between the groups were assessed using Student's *t*-test. A permutation-based FDR approach was applied to correct for multiple testing (FDR < 0.2 was considered significant).

Protein immunoprecipitation for interaction analysis

HEK293 cells were lysed in RIPA buffer supplemented with protease inhibitor cocktail. Lysates were homogenized by passing 10 times through a 27-gauge (27G) needle attached to a 1 mL syringe and centrifuged at 13 000 ***g*** for 15 min at 4 °C. After preclearing the supernatant with Protein A agarose beads (Pierce, Schwerte, Germany), the samples were incubated overnight with UBR5 antibody (Cell Signaling, #8755, 1 : 50) on the overhead shaker at 4 °C. Subsequently, samples were incubated with 30 μL of Protein A beads for 1 h at room temperature. After this incubation, samples were centrifuged 5 min at 5000 ***g*** and the pellet was washed three times with RIPA buffer. For elution of the proteins, the pellet was incubated with 2x Laemmli Buffer, boiled for 5 min and centrifuged 5 min at maximum speed. The supernatant was taken and loaded onto a SDS/PAGE gel for western blot analysis.

Western blot

Cells were scraped from tissue culture plates by cell scraping and lysed in protein cell lysis buffer (50 Mm Hepes pH 7.4, 150 Mm NaCl, 1 mM EDTA, 1% Triton X-100) supplemented with 2 mM sodium orthovanadate, 1 mM PMSF, and protease inhibitor mix). Lysates were homogenized through syringe needle (27G) followed by centrifugation at 8000 ***g*** for 5 min at 4 °C and then supernatants were collected. Protein concentrations were determined with a standard BCA protein assay (ThermoFisher Scientific). Approximately 30 μg of total protein was separated by SDS/PAGE, transferred to PVDF membranes (Millipore, Darmstadt, Germany) and subjected to immunoblotting. Western blot analysis was performed with anti-UBR5 (Cell Signaling, #8755, 1 : 1000), anti-FLAG (Sigma, F7425, 1 : 5000), anti-DKC1 (Abcam, Cambridge, UK, ab156877, 1 : 1000), anti-RPL7 (Genetex, Irvine, CA, USA, #114727, 1 : 500), anti-RPS27 (Proteintech, Manchester, UK, 15355-1-AP, 1 : 500), anti p53 (Proteintech, 10442-1-AP, 1 : 1000), and anti- β Actin (Abcam, ab8226, 1 : 10 000).

Immunocytochemistry

Cells were fixed with paraformaldehyde (4% in PBS) for 20 min, followed by permeabilization (0.2% Triton X-100 in PBS for 10 min) and blocking (3% BSA in 0.2% Triton X-100 in PBS for 10 min). Primary antibody was incubated for 1.5 h at room temperature [anti-FLAG (Sigma F7425, 1 : 100)] and incubated with secondary antibody [Alexa Fluor 488 Goat anti-Mouse (ThermoFisher Scientific, #A-11029, 1 : 500)], Alexa Fluor 568 F(ab')2 Fragment of Goat Anti-Rabbit IgG (H+L) (ThermoFisher Scientific, #A-21069, 1 : 500) ,and Hoechst 33342 (Life Technologies, Darmstadt, Germany, #1656104) for 1 h at room temperature. PBS and distilled water wash were followed before the cover slips were mounted on Mowiol (Sigma, #324590).

Ribosome fractionation

About 5% (w/v) to 50% (w/v) sucrose gradients were prepared from 5% and 50% sucrose solutions [20 mM Tris-

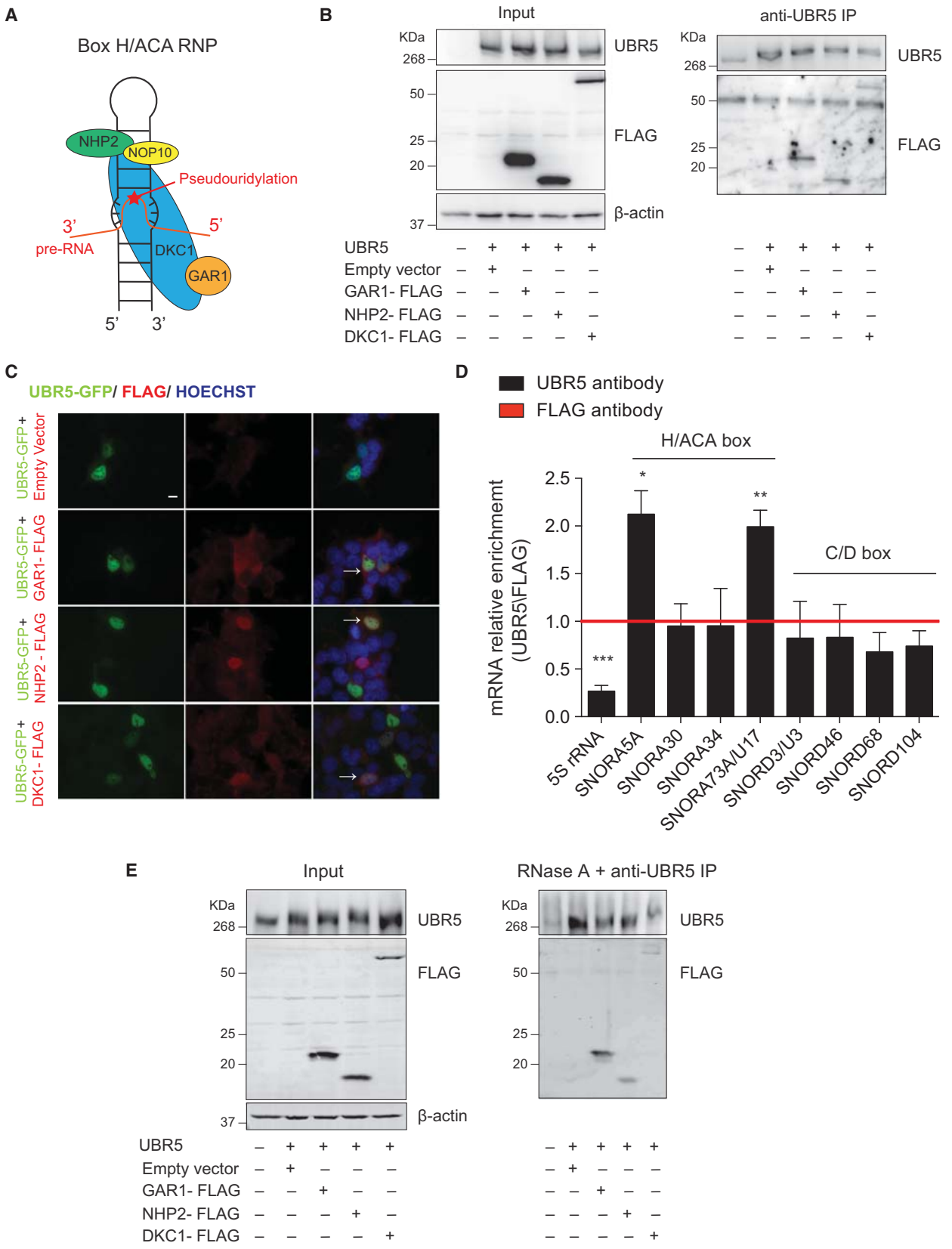

A
Box H/ACA RNP
NHP2
NOP10
Pseudouridylation
3'
5'
pre-RNA
DKC1
GAR1
5'
3'
B
Input
anti-UBR5 IP
KDa
268
50
25
20
37
UBR5
FLAG
β-actin
UBR5 – + + + +
Empty vector – + – – –
GAR1- FLAG – – + – –
NHP2- FLAG – – – + –
DKC1- FLAG – – – – +
C
UBR5-GFP/ FLAG/ HOECHST
UBR5-GFP + Empty Vector
UBR5-GFP + GAR1- FLAG
UBR5-GFP + NHP2 - FLAG
UBR5-GFP + DKC1- FLAG
D
UBR5 antibody
FLAG antibody
H/ACA box
C/D box
mRNA relative enrichmemt (UBR5\FLAG)
2.5
2.0
1.5
1.0
0.5
0.0
*
**
***
5S rRNA
SNORA5A
SNORA30
SNORA34
SNORA73A/U17
SNORD3/U3
SNORD46
SNORD68
SNORD104
E
Input
RNase A + anti-UBR5 IP
KDa
268
50
25
20
37
UBR5
FLAG
β-actin
UBR5 – + + + +
Empty vector – + – – –
GAR1- FLAG – – + – –
NHP2- FLAG – – – + –
DKC1- FLAG – – – – +

Fig. 1. UBR5 interacts with the H/ACA ribonucleoprotein (RNP) complex. (A) Schematic representation of the H/ACA RNP box with its protein components. (B) Immunoprecipitation with anti-UBR5 antibody in HEK293 overexpressing UBR5-GFP and GAR1-FLAG, NHP2-FLAG or DKC1-FLAG. Immunoprecipitation was followed by western blot with antibodies against UBR5 or FLAG (right panel). The input samples were loaded to assess overexpression levels and equal loading of total protein between samples (left panel). The images are representative of two independent experiments. (C) Immunocytochemistry of HEK293 cells with ectopic expression of UBR5-GFP and GAR1-FLAG, NHP2-FLAG or DKC1-FLAG. Hoechst staining was used as a marker of nuclei. Scale bar represents 10 μM. The images are representative of three independent experiments. (D) Ribonucleoprotein immunoprecipitation (RIP) with UBR5 antibody of HEK293T cells overexpressing GFP-UBR5. Quantitative PCR analysis of the indicated RNAs is expressed as fold enrichment over RIP performed with FLAG control antibody. Graph (relative enrichment to FLAG antibody) represents the mean ± SEM ($n = 3$). Statistical comparisons were made by Student's *t*-test for unpaired samples [*P*-value: *($P < 0.05$), **($P < 0.01$), ***($P < 0.001$)]. (E) Lysates of HEK293 overexpressing UBR5-GFP and GAR1-FLAG, NHP2-FLAG or DKC1-FLAG were treated with RNase A prior to immunoprecipitation with UBR5 antibody. Immunoprecipitation was followed by western blot with antibodies against UBR5 or FLAG. The images are representative of two independent experiments.

HCl (pH 7.4), 100 mM NaCl, 10 mM $MgCl_2$] supplemented with protease inhibitor (Roche, Berlin, Germany), 20 mM DTT (Sigma), and 0.1 mg·mL^{-1} cycloheximide (Sigma) in 11 mL ultracentrifuge polyallomer tubes (Beckman Coulter, Krefeld, Germany) using gradient maker (Biocomp, Munich, Germany). For fractionation, mESCs were lysed by incubation on ice 10 min with polysome extraction buffer (20 mM Tris-HCl (pH 7.4), 100 mM KCl, 10 mM $MgCl_2$, 1% Triton X-100) supplemented with 25 U·mL^{-1} DNase I, protease inhibitor, 20 mM DTT, and 0.1 mg·mL^{-1} cycloheximide followed by 1 h lysis at 4 °C. Lysates were centrifuged at 10 000 ***g*** for 10 min at 4 °C. The supernatant was collected, loaded on the gradient and centrifuged at 246 942 ***g*** for 3 h at 4 °C using SW-41 rotor and an ultracentrifuge (Beckman Coulter). The gradients were fractionated in equal volume while absorbance at 254 nm was recorded by fraction collector (Teledyne ISCO). Protein was extracted from each fraction by precipitating with Trichloroacetic acid (Sigma).

RNA immunoprecipitation-quantitative RT-PCR

RIP experiments with anti-UBR5 (Cell Signaling, #8755) and control anti-FLAG antibodies (Sigma, F7425) in HEK293T cells overexpressing GFP-UBR5 were performed following the protocol described in ref [21] with some modifications [22]. Cells were pelleted by centrifugation at 1000 ***g*** for 10 min at 4 °C and washed several times with ice cold PBS. The final cell pellet was resuspended with an equal volume of polysome lysis buffer [100 mM KCl, 5 mM $MgCl_2$, 10 mM HEPES (pH 7.0), 0.5% NP40 (Sigma)] supplemented with 1 mM DTT, 100 per units RNase Out (Invitrogen, Schwerte, Germany), 400 μM VRC (New England BioLabs, Frankfurt am Main, Germany), and protease inhibitor. Cell lysates were incubated on ice for 10 min and homogenized through syringe needle. Antibody coating of protein A beads was prepared by preswelling Protein A sepharose beads (Thermoscientific) in NT2 buffer (50 mM Tris–HCl (pH 7.4), 150 mM NaCl, 1 mM $MgCl_2$, 0.05% NP40) supplemented with 5% BSA to a final ratio of 1 : 5 for at least 1 h at 4 °C prior to use. Then, the antibody was added to bead slurry and incubated for 2 h at 4 °C. Immediately before use, antibody-coated beads were washed (five times) with ice cold NT2 buffer followed by resuspension of beads in ice cold NT2 buffer supplemented with 200 units of an RNase inhibitor, 400 μM vanadyl ribonucleoside complexes, 1 mM DTT, and 20 mM EDTA. The cell lysate was mixed with antibody-coated beads and incubated 2 h at room temperature. The beads were washed five times with ice cold NT2 buffer and finally washed with NT2 buffer with 1% Triton X-100. RNA extraction was done from the immunoprecipitated pellet using RNAbee (Tel-Test Inc., Friendswood, TX, USA). cDNA was generated using qScript cDNA SuperMix (Quantabio, Beverly, MA, USA). SybrGreen real-time qPCR experiments were performed with a 1 : 20 dilution of cDNA using a CFC384 Real-Time System (Bio-Rad) following the manufacturer's instructions. Data were analyzed with the comparative 2ΔΔCt method (RNA relative fold enrichment in anti-UBR5 RIP over RIP performed with control FLAG antibody) after normalization to the corresponding input values. See Table S1 for details about the primers used for this assay.

Bromodeoxyuridine proliferation assay

The mESCs were incubated with media containing 10 μM·mL^{-1} bromodeoxyuridine (BrdU) for the indicated times. Cells were fixed with formaldehyde 4% in PBS. Then, cells were permeabilized with 0.2% Triton X-100 in PBS for 10 min and blocked with 3% BSA-PBS for 1 h at room temperature. 2N HCl was added for 15 min at room temperature. After this, cells were incubated in 0.1 M sodium tetra-borate for 15 min at room temperature. We performed overnight incubation with rabbit anti-BrdU (ABD Serotech, OBT0030, 1 : 1000) at 4 °C followed by incubation with an anti-rat-AlexaFluor 546 Cross-Adsorbed (ThermoFisher Scientific, #A11081, 1 : 500) for 1 h. Hoechst 33342 was used to visualize nuclei.

RNA extraction and northern blotting

Total RNA was extracted using RNAbee (Tel-Test Inc.). About 2.5 μg of total RNA was resolved on a 1% agarose/

0.4 M formaldehyde gel using the tricine/triethanolamine buffer system (ref PMID 23800830). The RNA was blotted on an uncharged nylon membrane (Roth) by capillary transfer overnight in 10x SSC buffer. Following UV-crosslinking, the blots were prehybridised for 1 h in Church buffer. Hybridization with 5′-32P-labeled oligonucleotides in Church buffer was carried out overnight at 40 °C. The oligonucleotide sequences were as follows: ITS1 mouse: (5′-GCCGCTCCTCCACAGTCTCCCGTTTAATGATC-3′), ITS2 human: (5′-CTGCGAGGGAACCCCCAGCCGCG CA-3′), ITS2 mouse: (5′-CTGCGAGCAAACTCCCAGC CGCGCAGCAC-3′). The ITS1 (human) probe was generated by *in vitro* transcription with [α-32P]-GTP using the BamHI linearized pGEM4Z-ITS1 plasmid as a template. The signals were detected using a Typhoon FLA 7000 (GE Healthcare) and quantification was performed with Image Quant TL (GE Healthcare, Solingen, Germany).

Results

UBR5 interacts with the H/ACA ribonucleoprotein complex

UBR5 is upregulated in both hESCs and mESCs compared with their differentiated counterparts [15,16,20], suggesting a role for this protein in ESC function. To gain insights into the function of UBR5, we performed co-immunoprecipitation experiments of endogenous UBR5 in hESCs followed by single-shot proteomic analysis. We identified 101 interactors of UBR5 (Data S1). Gene Ontology (GO) analysis of UBR5 potential interactors indicated enrichment for proteins involved in biological processes such as spliceosome assembly, small nuclear RNA (snRNA) modification, and rRNA pseudouridine synthesis (Data S1 and Fig. S1). Interestingly, we found that UBR5 interacts with the proteins (i.e., DKC1, GAR1, NOP10, and NHP2) of the nucleolar H/ACA RNP complex (Data S1).

The H/ACA RNP box is formed by four different proteins bound to small nucleolar RNAs (snoRNAs) (Fig. 1A), and modulates the modification of rRNAs as well as other RNAs by catalyzing their pseudouridylation (i.e., conversion of uridine to pseudouridine) [23,24]. The noncoding H/ACA snoRNAs act as adaptors that link the catalytic protein DKC1 to its targets [25,26]. The three other proteins of the H/ACA RNP box – GAR1, NOP10, and NHP2 – act as scaffolding proteins and they are essential for the proper functioning of the complex [27,28]. We identified DKC1, GAR1, NOP10, and NHP2 as putative binding partners of UBR5 in our analysis, suggesting

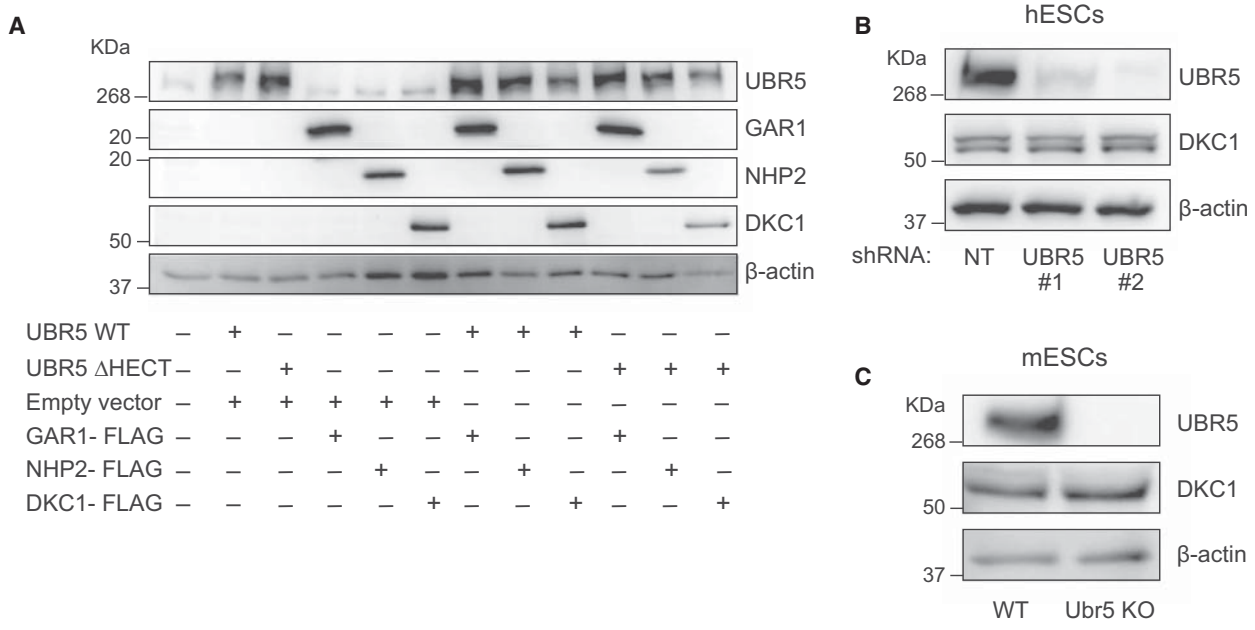


Fig. 2. UBR5 does not induce degradation of proteins of the H/ACA RNP complex. (A) Western blot analysis of $UBR5^{WT}$ or $UBR5^{\Delta HECT}$ and GAR1- FLAG, NHP2-FLAG or DKC1-FLAG using UBR5 and FLAG antibodies, respectively. GAR1, NHP2 AND DKC1 were overexpressed together with either $UBR5^{WT}$ or $UBR5^{\Delta HECT}$ in HEK293 cells. β-Actin is the loading control. The images are representative of three independent experiments. (B) Western blot of endogenous UBR5 and DKC1 in H9 hESC line upon UBR5 knockdown ($H9^{UBR5\ KD}$). β-Actin is the loading control. The images are representative of three independent experiments. (C) Western blot of UBR5 and DKC1 in mESCs upon UBR5 knockout ($mESC^{Ubr5\ KO}$). β-Actin is the loading control. The images are representative of two independent experiments.

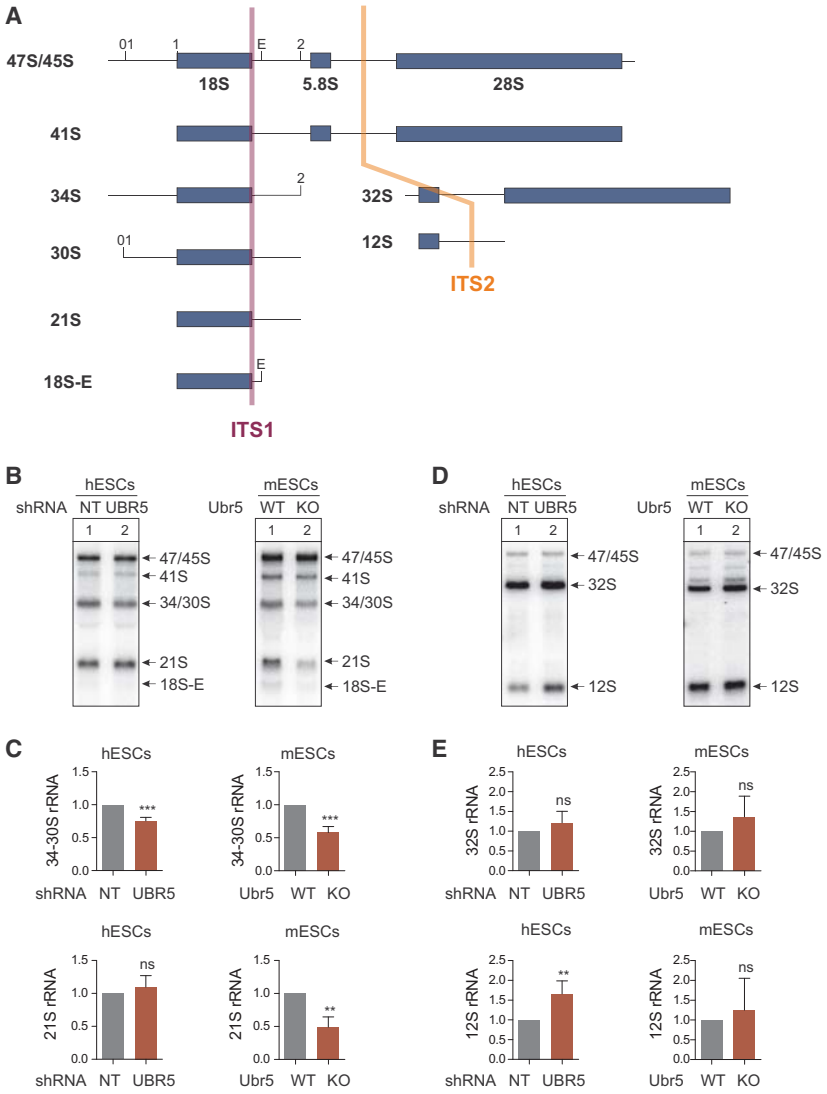


Fig. 3. UBR5 is required for the correct maturation of ribosomal RNA. (A) Schematic representation of the 47S/45S pre-rRNA processing pathway and the binding sites of the internal probes ITS1 and ITS2. The 47S pre-RNA is subjected to several internal and external cleavages to finally convert into the mature 18S, 5.8S and 28S rRNA. The intermediate species observed on the northern blots in (B)–(E) are listed. (B) Representative northern blots of total RNA obtained from hESCs upon *UBR5* knockdown (KD) and mESCs upon *Ubr5* knockout (KO). The hybridization was performed with probes targeting the ITS1 region of ribosomal precursor RNAs. NT: nontargeting shRNA control. (C) Quantification of the northern blots shown in (B). The amounts of the 30S/34S or 21S/20S precursor rRNAs were normalized to the corresponding amount of 47/45S pre-rRNA. The mean values ± SD were calculated from $n = 9$ (hESCs) and $n = 5$ (mESCs) independent experiments. Statistical comparisons were made by Student's *t*-test for paired samples [*P*-value: **($P < 0.01$), ***($P < 0.001$), ns: not significant]. (D) As performed in (B), but with a probe targeting the ITS2 region of ribosomal precursor RNAs. (E) Quantification of the blots shown in (D). The amounts of the 32S or 12S precursor rRNAs were normalized to the corresponding amount of 47/45S pre-rRNA. The mean values ± SD were calculated from $n = 6$ (hESCs) and $n = 4$ (mESCs). Statistical comparisons were made by Student's *t*-test for paired samples. [*P*-value: **($P < 0.01$), ns: not significant].

a role of UBR5 in the regulation of the H/ACA RNP complex (Data S1). To further assess this interaction, we overexpressed UBR5 together with GAR1, NHP2, or DKC1 in HEK293 human cells and performed co-immunoprecipitation assays followed by western blot. These experiments confirmed the interaction of UBR5 with GAR1, NHP2, and DKC1 (Fig. 1B). Prompted by these results, we analyzed the intracellular localization of UBR5 and GAR1, NHP2, or DKC1 in HEK293. Notably, these components of the H/ACA RNP complex co-localized with UBR5 in the nucleus, further supporting our interactome analysis (Fig. 1C).

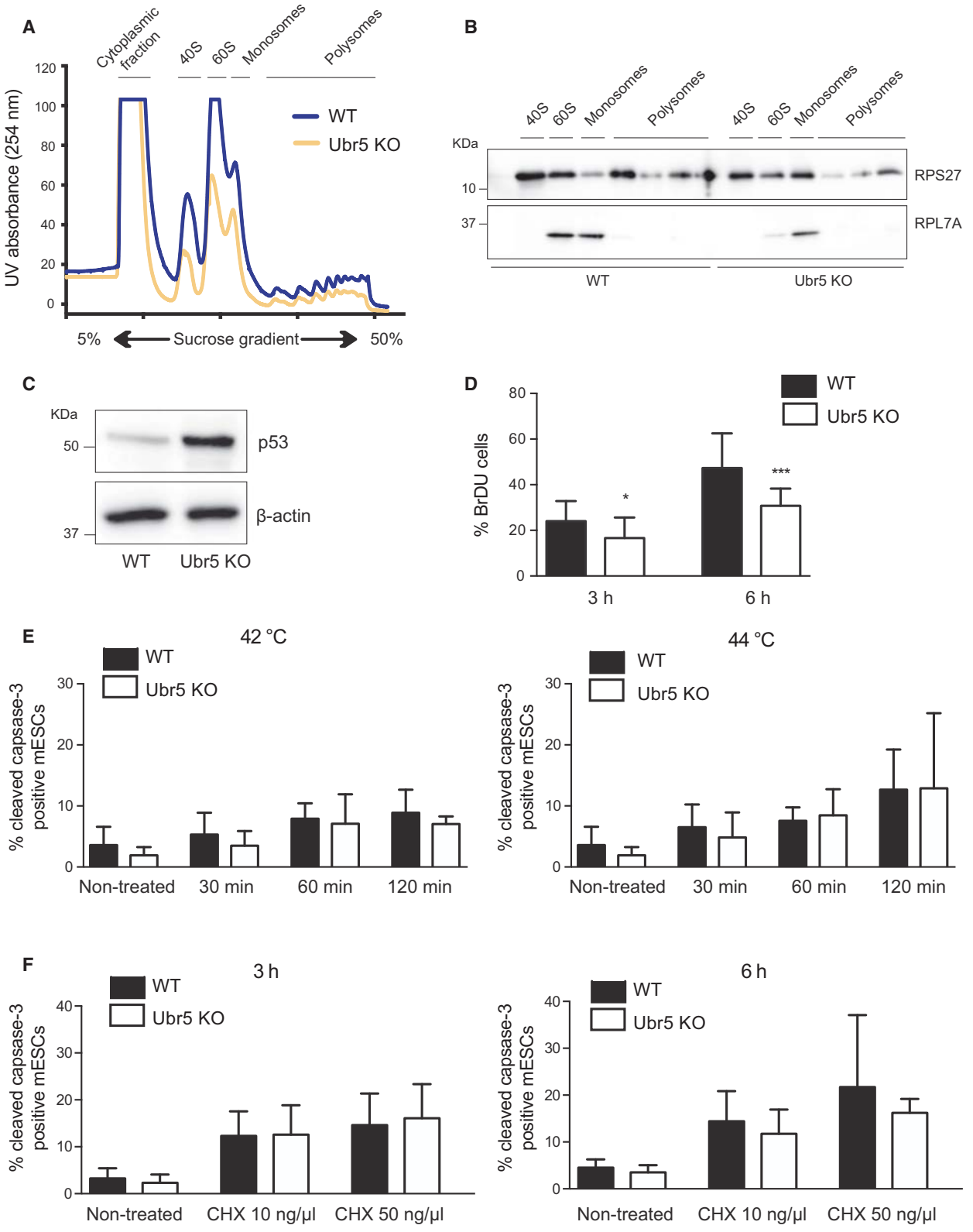

A
Cytoplasmic fraction
40S
60S
Monosomes
Polysomes
UV absorbance (254 nm)
WT
Ubr5 KO
5%
Sucrose gradient
50%
B
KDa
RPS27
RPL7A
WT
Ubr5 KO
C
KDa
p53
β-actin
WT
Ubr5 KO
D
% BrDU cells
3 h
6 h
E
42 °C
44 °C
% cleaved capsase-3 positive mESCs
Non-treated
30 min
60 min
120 min
F
3 h
6 h
CHX 10 ng/µl
CHX 50 ng/µl

Fig. 4. Loss of UBR5 impairs translational profile and triggers a stress response in mESCs. (A) Polysome profiling of $mESC^{WT}$ and $mESC^{Ubr5\ KO}$. Graph is representative of three experiments. (B) Western blot of the ribosomal fractions obtained in (A). Proteins were isolated from each fraction followed by western blot with antibodies against RPS27 (40S ribosomal subunit) and RPL7A (60S ribosomal subunit). The image is representative of two experiments. (C) p53 levels in $mESC^{WT}$ and $mESC^{Ubr5\ KO}$ cells. β-Actin is the loading control. The image is representative of four experiments. (D) BrDU-positive cells in $mESC^{WT}$ and $mESC^{Ubr5\ KO}$ after 3 h and 6 h incubation times. The graph represents mean ± SEM of at least 1200 cells scored in four different experiments. Statistical comparisons were made by Student's *t*-test for unpaired samples. [*P*-value: *($P < 0.05$), ***($P < 0.001$)]. (E) Percentage of cleaved caspase-3-positive cells in $mESC^{WT}$ and $mESC^{Ubr5\ KO}$ upon heat stress at 42 °C (left panel) and 44 °C (right panel) during the indicated periods. Graphs represent mean ± SEM of at least 700 cells scored in four different experiments. No significant differences were found. Statistical comparisons were made by Student's *t*-test for unpaired samples. (F) Percentage of cleaved caspase-3-positive cells upon cyclohexamide treatment for 3 h (left panel) and 6 h (right panel). Graphs represent mean ± SEM of at least 700 cells scored in four different experiments. No significant differences were found. Statistical comparisons were made by Student's *t*-test for unpaired samples.

With the strong interaction between UBR5 and proteins of the H/ACA RNP complex, we examined whether UBR5 protein also pulls down snoRNAs by RNA immunoprecipitation (RIP) assays. We assessed four snoRNAs of the H/ACA RNP complex and found interaction of UBR5 protein with SNORA73A/U17 and SNORA5A (Fig. 1D). In contrast, we did not observe significant interaction of UBR5 with snoRNAs of the C/D box RNP, a distinct nucleolar complex that catalyzes 2′ O-methylation of rRNA (Fig. 1D). Since UBR5 protein pulled down snoRNAs of the H/ACA RNP, we asked whether the interaction of UBR5 with proteins of this complex is mediated by RNA. To assess this hypothesis, we performed co-immunoprecipitation assays treating the samples with RNase A prior to immunoprecipitation with UBR5 antibody. Notably, the interactions with GAR1, NHP2, and DKC1 proteins remained upon RNase A treatment (Fig. 1E), indicating that this binding is not due to secondary, RNA-mediated interactions.

Given the role of UBR5 in polyubiquitination and proteasomal degradation of distinct substrates [15], we asked whether its interaction with H/ACA proteins triggers their degradation through the proteasome. For this purpose, we overexpressed GAR1, NHP2, or DKC1 with either wild-type UBR5 ($UBR5^{WT}$) or a catalytically dead mutant of UBR5 ($UBR5^{\Delta HECT}$) which lacks the ubiquitin ligase activity. We assessed the levels of DKC1, NHP2, and GAR1 by western blot and found no differences upon overexpression of either $UBR5^{WT}$ or $UBR5^{\Delta HECT}$, indicating that UBR5 does not target them for proteasomal degradation (Fig. 2A). In addition, we generated two independent hESC lines with reduced levels of UBR5 ($hESC^{UBR5\ KD}$) and found no changes in the endogenous levels of DKC1 (Fig. 2B). Likewise, endogenous DKC1 levels were not affected in a mESC line in which UBR5 was knockedout using CRISPR/Cas9 ($mESC^{Ubr5\ KO}$) (Fig. 2C). Thus, our results indicate that UBR5 interacts with distinct proteins of the H/ACA ribonucleoprotein complex but it does not induce their proteasomal degradation.

UBR5 is required for the correct maturation of ribosomal RNA

Given the interaction of UBR5 with the H/ACA RNP complex, we asked whether UBR5 has a role in the regulation of rRNA in ESCs. For this purpose, we assessed the maturation of rRNAs by northern blot analysis of their precursors in $hESC^{UBR5\ KD}$ and $mESC^{UBR5\ KO}$. In eukaryotes, the ribosome contains four distinct rRNAs: the 18S rRNA is present in the 40S small ribosomal subunit, while the large 60S subunit contains the 5.8S, 28S, and 5S rRNAs. The 18S, 5.8S, and 28S rRNA are transcribed by RNA polymerase I as a single rRNA precursor transcript, the 47S/45S pre-ribosomal RNA (pre-RNA) [29,30]. Ribosomal proteins and assembly factors bind co-transcriptionally to the pre-rRNA, which then undergoes a series of modifications and processing steps including endonucleolytic cleavage and exonucleolytic trimming [31]. The maturation process can be assessed by radioactive labeled probes (ITS1 and ITS2) which bind to internal sequences of the pre-RNA. These sequences flank the mature rRNA and are eliminated during the sequential cleavages of the pre-RNA [32] (Fig. 3A).

rRNA precursors of the 18S rRNA were detected with an ITS1 probe (Fig. 3B,C). In $hESC^{UBR5\ KD}$, the amount of 34/30S rRNA relative to the 47/45S pre-rRNA was significantly reduced compared with control hESCs. The 30S rRNA is further processed to 21S rRNA by additional cleavage steps at positions 01 and 1. Interestingly, the levels of 21S rRNA remained unchanged in $hESC^{UBR5\ KD}$ when compared with control hESCs. Since the 47S pre-rRNA can also be processed involving an alternative pathway via the 41S rRNA, the reduced cleavage into 30S rRNA could

potentially be compensated. In mESC$^{Ubr5\ KO}$, the reduction of 34/30S rRNA compared to 47/45S pre-rRNA was even more dramatic. Moreover, the levels of 20S rRNA were also affected in these cells. However, the amounts of 41S rRNA, the intermediate of the alternative pathway of cleavage at position 2, were unchanged. Using a probe targeting ITS2, we examined the 32S and 12S precursor rRNAs in hESC and mESCs (Fig. 3D,E). Although in hESCs$^{UBR5\ KD}$, the levels of 12S rRNA relative to 47/45S pre-rRNA were significantly increased, we did not observe these effects in mESC$^{Ubr5\ KO}$. In both hESCs and mESCs, the levels of 32S rRNA were not significantly changed upon loss of UBR5 (Fig. 3D,E). Taken together, our results indicate that dysfunction of UBR5 could affect rRNA maturation, particularly conversion into 34/30S rRNA.

UBR5 deficiency impairs translational profile and triggers a stress response in mESCs

Since alterations in rRNA might originate defects in the proper assembly of the ribosomes, we assessed ribosome pools in mESC$^{Ubr5\ KO}$ (Fig. 4A). When compared with wild-type ESCs, these cells exhibited a decrease in 40S and 60S, as well as in the monosome and polysome fraction, suggesting lower translational rates in the absence of UBR5. The levels of 40S (i.e., RPS27) and 60S (i.e., RPL7A) proteins in mESC$^{UBR5\ KO}$ cells were also altered in different ribosome fractions, although the decrease in RPL7A was more evident than RPS27 (Fig. 4B). Alterations in rRNA synthesis can trigger a cellular stress response [33,34], which, in turn, increases p53 levels. Notably, we found increased p53 levels in mESC$^{Ubr5\ KO}$, suggesting the activation of a stress response in these cells (Fig. 4C). p53-mediated response toward stress promotes cell cycle arrest to maintain cellular homeostasis [35]. Importantly, we observed that the proliferation rates in mESC$^{Ubr5\ KO}$ cells were significantly diminished (Fig. 4D). We then asked whether the lack of UBR5 sensitized mESCs toward other stressors which inhibit translation. For this purpose, we exposed mESC$^{Ubr5\ KO}$ cells to either heat stress (Fig. 4E) or cycloheximide (Fig. 4F) and determine the degree of apoptosis by measuring cleaved caspase-3 levels. However, we did not find notable differences in the apoptosis rates in the absence of UBR5. Altogether, our results suggest that UBR5 participates in the maturation of rRNA, a process necessary for proper ribosomal assembly. In the absence of UBR5, ESCs exhibit defects in their ribosomal content and increased p53-mediated stress response, resulting in lower proliferation rates.

Discussion

Growing evidence indicates that UBR5 is involved in the regulation of distinct RNA-related pathways [4,12,13]. In mESCs, UBR5 is required for the proper function of the miRNA machinery by directing the scaffolding protein GW182 to the miRNA complex through a ubiquitin ligase-independent mechanism [13]. This process results in the recruitment of other downstream effectors like DDX6 and Tob1/2, which promote the deadenylation and consequent destabilization of target mRNAs [36–38]. Moreover, UBR5 phosphorylation by p90 ribosomal S6 kinase (p90RSK) is necessary for the effect of UBR5 on miRNA repression activity, adding an additional regulatory level [39]. The effects of UBR5 on the miRNA pathway could be evolutionary conserved among species as similar effects were also reported in *Drosophila melanogaster* [40]. Additionally, upon UV-induced DNA damage, UBR5 represses the transcription at the damaged sites by inhibiting the elongation of polymerase II [41]. UBR5 also interacts with the phosphorylated form of PIH1D1, which is a subunit of the co-chaperone complex R2TP [42]. R2TP is implicated in the assembly of large complexes such as the H/ACA RNP complex [43]. However, a direct link between UBR5 and the H/ACA RNP complex has not been previously reported. Notably, we find an interaction between UBR5 and the four proteins of the nucleolar H/ACA complex (i.e., DKC1, GAR1, NOP10, and NHP2). Besides these proteins, the H/ACA RNP complex contains small nucleolar RNAs (snoRNAs) that complement rRNA sequences and guide their pseudouridylation, a modification required for proper rRNA processing. As a further indication of the link of UBR5 with H/ACA RNP, RIP experiments showed that UBR5 protein also pulls down snoRNAs of this complex while it does not interact with snoRNAs of the nucleolar C/D Box RNP. Importantly, our data suggest that the interaction of UBR5 with proteins of the H/ACA RNP complex is RNA-independent. Thus, these direct protein–protein interactions raise the interesting hypothesis that increased levels of UBR5 could titrate out proteins from the H/ACA RNP complex, leading to a lower stability of this complex. However, our results do not support this hypothesis because we observe that loss of UBR5 impairs the maturation of rRNA and increases the levels of intermediates, a process that could ensue from impaired functioning of the H/ACA RNP complex. In addition, we show that the levels of the 40S and 60S subunits of the ribosome, as well as the monosome and polysomes fractions, are reduced upon loss of UBR5, correlating with impaired

rRNA biogenesis. Taken together, these findings support that UBR5 is important for the function of the H/ACA RNP complex, but it does not act as a suppressor of its stability/activity. In support of this hypothesis, ESCs express intrinsic higher levels of UBR5 protein when compared with their differentiated counterparts, while they exhibit enhanced global translation rates [44].

Besides the nucleolar H/ACA RNP complex, small Cajal body-specific RNAs (scaRNAs) confined to Cajal bodies in the nucleus also contain similar box H/ACA domains as well as box C/D domains [45]. As such, scaRNAs are involved in the pseudouridylation or 2′ O-methylation required for maturation of not only small nuclear RNAs (snRNAs) but also nucleolar snoRNAs [45]. Since Cajal bodies contribute to the biogenesis of nucleolar RNPs [45], it will be fascinating to examine whether UBR5 interacts with nuclear RNPs in Cajal bodies and the impact of this interaction in nucleolar rRNA metabolism. Besides its effects in nucleolar RNP biogenesis, Cajal bodies are also involved in the proper maturation of snRNAs required for the assembly of distinct nuclear RNP complexes such as the spliceosome, which regulates RNA splicing [46]. Notably, the interactome of UBR5 is highly enriched in proteins involved in the assembly and activity of the spliceosome RNP complex (Fig. S1 and Data S1). Since these data support an interaction between UBR5 and spliceosome, it will be interesting to assess a role of UBR5 in the assembly and/or activity of the spliceosome RNP complex.

The link of UBR5 with rRNA biogenesis indicates a role in translational rates. Importantly, UBR5 contains a PABC domain at the C terminus, also known as MLLE domain, which mediates protein–protein interaction through the binding of the (PABP-interacting motifs) PAM2 peptide motif [47,48]. This PABC domain is equivalent to the one found in the poly(A) binding protein (PABPC1), a protein that recognizes 3′ mRNA poly(A) tails and plays an essential role in eukaryotic translation initiation [49,50]. UBR5 and PABPC1 share common interacting partners, such as Paip1 [47] and Paip2 [51]. Paip2 is a repressor of translation initiation which can interact with both PABPC1 and UBR5. When bound to UBR5, Paip2 is ubiquitinated and degraded by the proteasome [51]. Conversely, the levels of the repressor increase in the absence of UBR5 and translation are consequently inhibited. Thus, UBR5 could modulate translational rates via both Paip2-mediated mechanisms and the maturation of rRNA.

As a consequence of perturbations in ribosomal biogenesis, cells enter into the so called 'nucleolar stress' [52]. Under normal conditions, the E3 ligase Mdm2 interacts with p53 and results in the ubiquitination and degradation of p53. During nucleolar stress, several assembly factors, such as Arf or Nucleophosmin1, and ribosomal proteins bind to Mdm2 preventing its interaction with p53, a process that results in stabilization of p53 levels [53,54]. In addition, the 5S rRNA also interacts and blocks Mdm2 [55]. As a consequence of p53 accumulation, cell enters in cell cycle arrest to ensure cellular homeostasis until the stress is over. Remarkably, we find that the levels of p53 levels are increased in mES-$C^{UBR5\ KO}$ cells. This could indicate that UBR5-lacking cells undergo nucleolar stress, as suggested by the altered rRNA precursors and polysome levels. Moreover, loss of UBR5 induces a decline in the proliferation rates of ESCs, which could be ensue from increased p53 levels. In these lines, UBR5 is deregulated in many cancer types and alters p53 levels [56–58]. However, it is unclear whether UBR5 would act as a tumor suppressor or oncogene, since both amplifications as well as loss-of-function mutations have been linked to tumorigenesis. Thus, the cellular context might be determinant for the role of UBR5 in p53 regulation and cellular proliferation [59–61]. Whereas previous studies reported that the role of UBR5 in the regulation of mRNA machinery is necessary for cell proliferation [13,62], here we provide data indicating that UBR5 also modulates cell proliferation via modulation of rRNA maturation. Thus, UBR5 could impinge upon cellular proliferation through multiple pathways.

Taken together, our results indicate that UBR5 participates in the maturation of rRNA through regulating the nucleolar H/ACA RNP complex. This regulatory pathway is required for proper rRNA processing and ribosomal assembly. Upon loss of UBR5, rRNA intermediates accumulate, decreasing the pool of ribosomes. This perturbation may activate the nucleolar stress response, triggering the accumulation of p53 and consequent proliferation arrest. Taking into consideration the multiple roles of UBR5 in cell cycle and cellular homeostasis, our findings add an additional regulation layer to this process.

Acknowledgements

This work was supported by The Deutsche Forschungsgemeinschaft (DFG) (CECAD and VI742/1-2). We thank Volker Böhm for helping with data analysis.

Author contributions

IS and DV conceived and supervised the study. IS performed most of the experiments, data analysis and

interpretation through discussions with NHG and DV. JVG assessed rRNA maturation. SK performed western blot experiments. HJL performed RIP experiments. MH, VK and MSD generated mESCubr5 KO line and contributed to experiment design. NHG and CD contributed with their expertise on RNA metabolism and performed data analysis. IS, JVG, NHG and DV wrote the manuscript. All the authors discussed the results and commented on the manuscript.

Data Accessibility

Research data pertaining to this article is located at figshare.com: https://dx.doi.org/10.6084/m9.figshare.9153215.

References

1 Callaghan MJ, Russell AJ, Woollatt E, Sutherland GR, Sutherland RL and Watts CK (1998) Identification of a human HECT family protein with homology to the Drosophila tumor suppressor gene hyperplastic discs. *Oncogene* **17**, 3479–3491.

2 Henderson MJ, Russell AJ, Hird S, Muñoz M, Clancy JL, Lehrbach GM, Calanni ST, Jans DA, Sutherland RL and Watts CK (2002) EDD, the human hyperplastic discs protein, has a role in progesterone receptor coactivation and potential involvement in DNA damage response. *J Biol Chem* **277**, 26468–26478.

3 Tasaki T, Mulder LC, Iwamatsu A, Lee MJ, Davydov IV, Varshavsky A, Muesing M and Kwon YT (2005) A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol* **25**, 7120–7136.

4 Kozlov G, Trempe JF, Khaleghpour K, Kahvejian A, Ekiel I and Gehring K (2001) Structure and function of the C-terminal PABC domain of human poly(A)-binding protein. *Proc Natl Acad Sci USA* **98**, 4409–4413.

5 Henderson MJ, Munoz MA, Saunders DN, Clancy JL, Russell AJ, Williams B, Pappin D, Khanna KK, Jackson SP, Sutherland RL *et al.* (2006) EDD mediates DNA damage-induced activation of CHK2. *J Biol Chem* **281**, 39990–40000.

6 Munoz-Escobar J, Matta-Camacho E, Kozlov G and Gehring K (2015) The MLLE domain of the Ubiquitin Ligase UBR5 binds to its catalytic domain to regulate substrate binding. *J Biol Chem* **290**, 22841–22850.

7 Rutz S, Kayagaki N, Phung QT, Eidenschenk C, Noubade R, Wang X, Lesch J, Lu R, Newton K, Huang OW *et al.* (2015) Deubiquitinase DUBA is a post-translational brake on interleukin-17 production in T cells. *Nature* **518**, 417–421.

8 Eblen ST, Kumar NV, Shah K, Henderson MJ, Watts CK, Shokat KM and Weber MJ (2003) Identification of novel ERK2 substrates through use of an engineered kinase and ATP analogs. *J Biol Chem* **278**, 14926–14935.

9 Munoz MA, Saunders DN, Henderson MJ, Clancy JL, Russell AJ, Lehrbach G, Musgrove EA, Watts CK and Sutherland RL (2007) The E3 ubiquitin ligase EDD regulates S-phase and G(2)/M DNA damage checkpoints. *Cell Cycle* **6**, 3070–3077.

10 Mu JJ, Wang Y, Luo H, Leng M, Zhang J, Yang T, Besusso D, Jung SY and Qin J (2007) A proteomic analysis of ataxia telangiectasia-mutated (ATM)/ATM-Rad3-related (ATR) substrates identifies the ubiquitin-proteasome system as a regulator for DNA damage checkpoints. *J Biol Chem* **282**, 17330–17334.

11 Kim MA, Kim HJ, Brown AL, Lee MY, Bae YS, Park JI, Kwak JY, Chung JH and Yun J (2007) Identification of novel substrates for human checkpoint kinase Chk1 and Chk2 through genome-wide screening using a consensus Chk phosphorylation motif. *Exp Mol Med* **39**, 205–212.

12 Cojocaru M, Bouchard A, Cloutier P, Cooper JJ, Varzavand K, Price DH and Coulombe B (2011) Transcription factor IIS cooperates with the E3 ligase UBR5 to ubiquitinate the CDK9 subunit of the positive transcription elongation factor B. *J Biol Chem* **286**, 5012–5022.

13 Su H, Meng S, Lu Y, Trombly MI, Chen J, Lin C, Turk A and Wang X (2011) Mammalian hyperplastic discs homolog EDD regulates miRNA-mediated gene silencing. *Mol Cell* **43**, 97–109.

14 Saunders DN, Hird SL, Withington SL, Dunwoodie SL, Henderson MJ, Biben C, Sutherland RL, Ormandy CJ and Watts CK (2004) Edd, the murine hyperplastic disc gene, is essential for yolk sac vascularization and chorioallantoic fusion. *Mol Cell Biol* **24**, 7225–7234.

15 Koyuncu S, Saez I, Lee HJ, Gutierrez-Garcia R, Pokrzywa W, Fatima A, Hoppe T and Vilchez D (2018) The ubiquitin ligase UBR5 suppresses proteostasis collapse in pluripotent stem cells from Huntington's disease patients. *Nat Commun* **9**, 2886.

16 Buckley SM, Aranda-Orgilles B, Strikoudis A, Apostolou E, Loizou E, Moran-Crusio K, Farnsworth CL, Koller AA, Dasgupta R, Silva JC *et al.* (2012) Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell Stem Cell* **11**, 783–798.

17 Elling U, Taubenschmid J, Wirnsberger G, O'Malley R, Demers SP, Vanhaelen Q, Shukalyuk AI, Schmauss G, Schramek D, Schnuetgen F *et al.* (2011) Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* **9**, 563–574.

18 Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA and Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–2308.
19 Gudjonsson T, Altmeyer M, Savic V, Toledo L, Dinant C, Grøfte M, Bartkova J, Poulsen M, Oka Y, Bekker-Jensen S *et al.* (2012) TRIP12 and UBR5 suppress spreading of chromatin ubiquitylation at damaged chromosomes. *Cell* **150**, 697–709.
20 Saez I, Koyuncu S, Gutierrez-Garcia R, Dieterich C and Vilchez D (2018) Insights into the ubiquitin-proteasome system of human embryonic stem cells. *Sci Rep* **8**, 4092.
21 Keene JD, Komisarow JM and Friedersdorf MB (2006) RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat Protoc* **1**, 302–307.
22 Lee HJ, Bartsch D, Xiao C, Guerrero S, Ahuja G, Schindler C, Moresco JJ, Yates JR 3rd, Gebauer F, Bazzi H *et al.* (2017) A post-transcriptional program coordinated by CSDE1 prevents intrinsic neural differentiation of human embryonic stem cells. *Nat Commun* **8**, 1456.
23 Matera AG, Terns RM and Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8**, 209–220.
24 Hamma T and Ferre-D'Amare AR (2010) The box H/ACA ribonucleoprotein complex: interplay of RNA and protein structures in post-transcriptional RNA modification. *J Biol Chem* **285**, 805–809.
25 Baker DL, Youssef OA, Chastkofsky MI, Dy DA, Terns RM and Terns MP (2005) RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. *Genes Dev* **19**, 1238–1248.
26 Charpentier B, Muller S and Branlant C (2005) Reconstitution of archaeal H/ACA small ribonucleoprotein complexes active in pseudouridylation. *Nucleic Acids Res* **33**, 3133–3144.
27 Hamma T, Reichow SL, Varani G and Ferre-D'Amare AR (2005) The Cbf5-Nop10 complex is a molecular bracket that organizes box H/ACA RNPs. *Nat Struct Mol Biol* **12**, 1101–1107.
28 Rashid R, Liang B, Baker DL, Youssef OA, He Y, Phipps K, Terns RM, Terns MP and Li H (2006) Crystal structure of a Cbf5-Nop10-Gar1 complex and implications in RNA-guided pseudouridylation and dyskeratosis congenita. *Mol Cell* **21**, 249–260.
29 Fromont-Racine M, Senger B, Saveanu C and Fasiolo F (2003) Ribosome assembly in eukaryotes. *Gene* **313**, 17–42.
30 Fatica A and Tollervey D (2002) Making ribosomes. *Curr Opin Cell Biol* **14**, 313–318.
31 Henras AK, Soudet J, Gérus M, Lebaron S, Caizergues-Ferrer M, Mougin A and Henry Y (2008) The post-transcriptional steps of eukaryotic ribosome biogenesis. *Cell Mol Life Sci* **65**, 2334–2359.
32 Granneman S and Baserga SJ (2004) Ribosome biogenesis: of knobs and RNA processing. *Exp Cell Res* **296**, 43–50.
33 Donati G, Montanaro L and Derenzini M (2012) Ribosome biogenesis and control of cell proliferation: p53 is not alone. *Cancer Res* **72**, 1602–1607.
34 Deisenroth C and Zhang Y (2010) Ribosome biogenesis surveillance: probing the ribosomal protein-Mdm2-p53 pathway. *Oncogene* **29**, 4253–4260.
35 Vogelstein B, Lane D and Levine AJ (2000) Surfing the p53 network. *Nature* **408**, 307–310.
36 Ezzeddine N, Chang TC, Zhu W, Yamashita A, Chen CY, Zhong Z, Yamashita Y, Zheng D and Shyu AB (2007) Human TOB, an antiproliferative transcription factor, is a poly(A)-binding protein-dependent positive regulator of cytoplasmic mRNA deadenylation. *Mol Cell Biol* **27**, 7791–7801.
37 Fabian MR, Mathonnet G, Sundermeier T, Mathys H, Zipprich JT, Svitkin YV, Rivas F, Jinek M, Wohlschlegel J, Doudna JA *et al.* (2009) Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol Cell* **35**, 868–880.
38 Zekri L, Huntzinger E, Heimstadt S and Izaurralde E (2009) The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. *Mol Cell Biol* **29**, 6220–6231.
39 Cho JH, Kim SA, Seo YS, Park SG, Park BC, Kim JH and Kim S (2017) The p90 ribosomal S6 kinase-UBR5 pathway controls Toll-like receptor signaling via miRNA-induced translational inhibition of tumor necrosis factor receptor-associated factor 3. *J Biol Chem* **292**, 11804–11814.
40 Zhou R, Hotta I, Denli AM, Hong P, Perrimon N and Hannon GJ (2008) Comparative analysis of argonaute-dependent small RNA pathways in Drosophila. *Mol Cell* **32**, 592–599.
41 Sanchez A, De Vivo A, Uprety N, Kim J, Stevens SM Jr and Kee Y (2016) BMI1-UBR5 axis regulates transcriptional repression at damaged chromatin. *Proc Natl Acad Sci USA* **113**, 11243–11248.
42 Horejsi Z, Stach L, Flower TG, Joshi D, Flynn H, Skehel JM, O'Reilly NJ, Ogrodowicz RW, Smerdon SJ and Boulton SJ (2014) Phosphorylation-dependent PIH1D1 interactions define substrate specificity of the R2TP cochaperone complex. *Cell Rep* **7**, 19–26.
43 Machado-Pinilla R, Liger D, Leulliot N and Meier UT (2012) Mechanism of the AAA+ ATPases pontin and reptin in the biogenesis of H/ACA RNPs. *RNA* **18**, 1833–1845.
44 You KT, Park J and Kim VN (2015) Role of the small subunit processome in the maintenance of pluripotent stem cells. *Genes Dev* **29**, 2004–2009.

45 Meier UT (2017) RNA modification in Cajal bodies. *RNA Biol* **14**, 693–700.
46 Cioce M and Lamond AI (2005) Cajal bodies: a long history of discovery. *Annu Rev Cell Dev Biol* **21**, 105–131.
47 Deo RC, Sonenberg N and Burley SK (2001) X-ray structure of the human hyperplastic discs protein: an ortholog of the C-terminal domain of poly(A)-binding protein. *Proc Natl Acad Sci USA* **98**, 4414–4419.
48 Kozlov G, De Crescenzo G, Lim NS, Siddiqui N, Fantus D, Kahvejian A, Trempe JF, Elias D, Ekiel I, Sonenberg N *et al*. (2004) Structural basis of ligand recognition by PABC, a highly specific peptide-binding domain found in poly(A)-binding protein and a HECT ubiquitin ligase. *EMBO J* **23**, 272–281.
49 Kahvejian A, Svitkin YV, Sukarieh R, M'Boutchou MN and Sonenberg N (2005) Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev* **19**, 104–113.
50 Mangus DA, Evans MC and Jacobson A (2003) Poly (A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol* **4**, 223.
51 Yoshida M, Yoshida K, Kozlov G, Lim NS, De Crescenzo G, Pang Z, Berlanga JJ, Kahvejian A, Gehring K, Wing SS *et al*. (2006) Poly(A) binding protein (PABP) homeostasis is mediated by the stability of its inhibitor, Paip2. *EMBO J* **25**, 1934–1944.
52 James A, Wang Y, Raje H, Rosby R and DiMario P (2014) Nucleolar stress with and without p53. *Nucleus* **5**, 402–426.
53 Zhang Y and Lu H (2009) Signaling to p53: ribosomal proteins find their way. *Cancer Cell* **16**, 369–377.
54 Chen D, Zhang Z, Li M, Wang W, Li Y, Rayburn ER, Hill DL, Wang H and Zhang R (2007) Ribosomal protein S7 as a novel modulator of p53-MDM2 interaction: binding to MDM2, stabilization of p53 protein, and activation of p53 function. *Oncogene* **26**, 5029–5037.
55 Nishimura K, Kumazawa T, Kuroda T, Katagiri N, Tsuchiya M, Goto N, Furumai R, Murayama A, Yanagisawa J and Kimura K (2015) Perturbation of ribosome biogenesis drives cells into senescence through 5S RNP-mediated p53 activation. *Cell Rep* **10**, 1310–1323.
56 Shearer RF, Iconomou M, Watts CK and Saunders DN (2015) Functional roles of the E3 Ubiquitin Ligase UBR5 in cancer. *Mol Cancer Res* **13**, 1523–1532.
57 Ling S and Lin WC (2011) EDD inhibits ATM-mediated phosphorylation of p53. *J Biol Chem* **286**, 14972–14982.
58 Smits VA (2012) EDD induces cell cycle arrest by increasing p53 levels. *Cell Cycle* **11**, 715–720.
59 Tomaic V, Pim D, Thomas M, Massimi P, Myers MP and Banks L (2011) Regulation of the human papillomavirus type 18 E6/E6AP ubiquitin ligase complex by the HECT domain-containing protein EDD. *J Virol* **85**, 3120–3127.
60 Ong SS, Goktug AN, Elias A, Wu J, Saunders D and Chen T (2014) Stability of the human pregnane X receptor is regulated by E3 ligase UBR5 and serine/threonine kinase DYRK2. *Biochem J* **459**, 193–203.
61 Matsuura K, Huang NJ, Cocce K, Zhang L and Kornbluth S (2017) Downregulation of the proapoptotic protein MOAP-1 by the UBR5 ubiquitin ligase and its role in ovarian cancer resistance to cisplatin. *Oncogene* **36**, 1698–1706.
62 Plank M, Hu G, Silva AS, Wood SH, Hesketh EE, Janssens G, Macedo A, de Magalhães JP and Church GM (2013) An analysis and validation pipeline for large-scale RNAi-based screens. *Sci Rep* **3**, 1076.

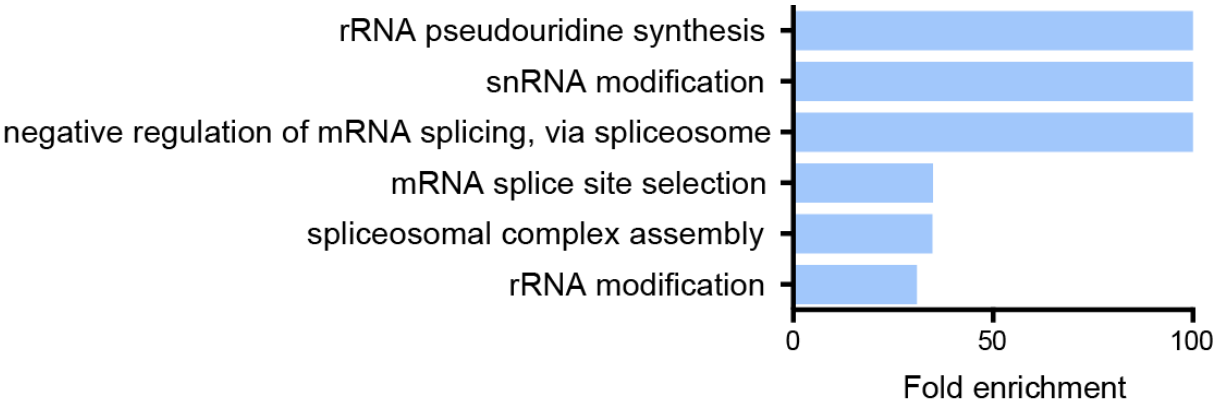
Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Gene Ontology analysis of UBR5 potential interactors indicated enrichment for proteins involved in RNA metabolism.

Table S1. List of primers used for qPCR.

Data S1. Analysis of proteomics data from co-immunoprecipitation experiments with UBR5 and FLAG antibodies in hESCs (FDR (q-value) <0.2 was considered significant).



Supplementary Figure 1. Gene Ontology analysis of UBR5 potential interactors indicated enrichment for proteins involved in RNA metabolism. Bar graph representing the top Gene Ontology Biological Processes enriched in the UBR5 interactome ($P < 0.05$). Analysis was performed using PANTHER software and Gene Ontology Consortium.

Gene	Forward (5' → 3')	Reverse (5' → 3')
5S rRNA	GGCCATACCACCCTGAACGC	CAGCACCCGGTATTCCCAGG
SNORA5A	CAGTACCTGTCCTATGCATGG	ACCAAATTTATCCCTGAGCCT
SNORA30	GCACTTTCACAGTTCCTTCC	CAGGGCAAGAATACAATCAAGG
SNORA34	CTGAAGACCAGCAGTTGTACT	TGAAATAGCCATTCCCTACTGAG
SNORA73A/U17	ATACACCCGGGAGGTCACTC	TGTTTCCTGCATGGTTTGTC
SNORD3/U3	AGAGGTAGCGTTTTCTCCTGAGCG	ACCACTCAGACCGCGTTCTC
SNORD46	AAGAATCCTTAGGCGTGGTTG	GTAACTATGACAAGTCCTTGCATTG
SNORD68	CGTGATGACATTCTCCGGAA	AATCAGATGGAAAAGGGTTCA
SNORD104	CCTGCTGTGATGACATTCCA	CAGGCTCAGACTCCAGTTC

Supplementary Table 1. List of primers used for qPCR.

Biochemical Society Transactions (2018) **46** 153–161
https://doi.org/10.1042/BST20170059

Review Article

The exon junction complex: structural insights into a faithful companion of mammalian mRNPs

Jennifer V. Gerbracht and Niels H. Gehring

Institute for Genetics, Department of Biology, University of Cologne, Zuelpicher Str. 47a, 50674 Cologne, Germany

Correspondence: Niels H. Gehring (ngehring@uni-koeln.de)

During splicing, the exon junction complex (EJC) is deposited upstream of exon-exon boundaries. The EJC and its peripheral bound proteins play an essential role in mediating mRNA export, translation and turnover. However, the exact sequence of EJC assembly and the involved factors during splicing remain elusive. Recently published structures of the human C* spliceosome clarified the position of the EJC at this phase of splicing and have given insight into previously unidentified interactions between the EJC and spliceosomal proteins. Here, these new observations are presented and the significance for EJC assembly is discussed. Furthermore, the vast landscape of EJC interacting proteins and their manifold functions are described. Finally, the factors involved in EJC disassembly and recycling are recapitulated. This review aims to integrate structural, biochemical and physiological data to obtain a comprehensive picture of EJC components during the lifetime of the EJC.

Introduction

RNA-binding proteins are an integral part of messenger ribonucleoproteins and determine the fate of mRNAs in the cell [1]. A prominent example is the exon junction complex (EJC) which is deposited at exon-exon junctions of mRNAs during splicing. The EJC remains bound to the mature mRNA until the first round of translation in the cytoplasm, when it is dissociated from the mRNA and disassembled [2,3]. During the lifetime of a mature mRNA, the EJC recruits peripheral proteins that are involved in processes such as pre-mRNA splicing, mRNA export, translation and nonsense-mediated mRNA decay (NMD), reviewed in [4,5]. The EJC therefore serves as a mark that connects splicing with post-transcriptional regulation of mRNA [5].

The significance of splicing exceeds its basic function that is the removal of introns and subsequent ligation of exons during the maturation of pre-mRNA. For example, alternative splicing is a regulated process which is involved in the control of tissue- and development-specific expression of splicing isoforms [6]. The splicing reaction is carried out by the spliceosome, a large ribonucleoprotein complex [7]. The spliceosome consists of five small nuclear RNA molecules with more than 100 associated proteins, which assemble by stepwise binding to the pre-mRNA and catalyze two transesterification steps. Substantial structural rearrangements and compositional changes occur during splicing, which have been grouped into characteristic stages (reviewed in ref [8]). Recently, the structures of the human B (pre-catalytic spliceosome) and C* (catalytically activated before second step) complexes have been solved by cryo-electron microscopy (cryo-EM) [9–11]. Besides providing valuable insight into the splicing process, these recent structural advancements have shed light on the molecular positioning of the EJC and interacting proteins. In this review, these recent findings are discussed and put into context of our understanding of EJC assembly and disassembly.

Received: 6 November 2017
Revised: 13 December 2017
Accepted: 18 December 2017

Version of Record published:
19 January 2018

Biochemical Society Transactions (2018) **46** 153–161
https://doi.org/10.1042/BST20170059

The core of the EJC

The EJC is a multiprotein complex formed by four core proteins [12,13]. The DEAD-box protein EIF4A3 (eukaryotic initiation factor 4A3) represents the RNA-binding platform of the EJC [14]. MAGOH and RBM8A (RNA-binding motif 8A, also known as Y14) form a very stable heterodimer interacting with RNA-bound EIF4A3 [12,13]. In mammals, two MAGOH paralogues exist (MAGOH and MAGOHB), which can structurally and functionally substitute each other [15]. CASC3 (also known as metastatic lymph node 51, MLN51, or Barentsz, BTZ), the fourth component of the EJC core, directly binds to EIF4A3 via its SELOR domain [16]. Interestingly, CASC3 is a shuttling protein that resides mainly in the cytoplasm, whereas the other core proteins are predominantly localized in the nucleus [16,17]. Developmental studies in *Drosophila* and mouse models suggest a distinct role of CASC3 in comparison with the other EJC components: While organisms lacking EIF4A3, MAGOH or RBM8A display severe general developmental impairments, the function of CASC3 in development seems to be limited to specific cells and tissues [18–20].

Reconstitution studies *in vitro* have provided valuable insights into the mechanism of EJC assembly and RNA binding [14]. In the presence of ATP, EIF4A3 adopts a closed conformation, in which it binds to the phosphate backbone of RNA [12,13]. Since the RNA-binding surface of EIF4A3 does not contact any of the RNA bases, its RNA binding is sequence-independent [12,13]. MAGOH-RBM8A locks EIF4A3 in its closed conformation, prevents its conversion to the open conformation and thereby stabilizes EIF4A3 binding to mRNA [21]. This mechanism ensures that the EJC remains bound to its target mRNA until it is actively disassembled.

Structural details about EJC assembly

In contrast to the spontaneous association of the four EJC components with RNA *in vitro*, their deposition *in vivo* is strictly splicing-dependent and occurs mainly at a position ∼20–24 nucleotides (nt) upstream of spliced exon-exon junctions, shown in Figure 1 [22]. This dependency on splicing can be explained by the fact that EIF4A3 directly interacts with the spliceosomal protein CWC22 [23–25]. CWC22 is an essential splicing factor that joins the activated spliceosomal B^{act} complex [26,27]. The two central domains indispensable for CWC22's splicing function are the MA3 and the middle of initiation factor 4G (MIF4G) domain, the latter of which interacts with EIF4A3 (Table 1) [23–25].

In the structures of the human C* spliceosome, the MA3 and MIF4G domains of CWC22 flank the 5′ exon of the mRNA [9,10]. As expected, MIF4G is closely associated with EIF4A3 (Figure 2A). EIF4A3 is localized 20–24 nts upstream of the exon–intron boundary (Figure 1), where it binds six nucleotides of the 5′ exon [9]. This position is consistent with the canonical EJC binding site after the completion of splicing. The interaction of EIF4A3 with the spliceosome is reinforced by the interaction of its N-terminal RecA1 domain with the spliceosomal protein EFTUD2 (Figure 2B). Interestingly, this part of EIF4A3 also interacts with the core protein CASC3 in atomic structures of the tetrameric EJC [12,13]. In the isolated EJC structure, two regions of CASC3's SELOR domain separately interact with the two RecA domains of EIF4A3. The N-terminal stretch 168–196 contacts the RecA2 domain, while the 214–245 stretch of CASC3 is moulded on the RecA1 domain. Both regions are connected via a flexible linker. In the C* spliceosome, however, only the fragment 170–194 of CASC3 is present and interacts with RecA2 close to the 5′ exon (Figure 1).

The MIF4G domain of CWC22 not only recruits EIF4A3 to the spliceosome, but also interacts with the core protein RBM8A and with part of the catalytic core of the spliceosome, the GTPase EFTUD2/Snu114. The MA3 domain of CWC22 contacts PRP8, the spliceosome's catalytic subunit [9,10], which might explain why this domain is important for the splicing function of CWC22 [27]. An additional CWC22-interacting partner present in the C* spliceosome is the human step II factor SLU7. The N-terminal part of SLU7 forms a tripod-like structure and is believed to stabilize the binding of the 5′ exon to the spliceosome [9,10]. While SLU7 mainly interacts with PRP8, one leg contacts the MA3 domain and another leg interacts with the MIF4G domain of CWC22. Thus, CWC22 contacts two of the EJC core components and is directly connected to proteins required for structural integrity and function of the spliceosome. Whether these spliceosomal factors aid in the recruitment of CWC22 and the EJC or rather stabilize the position of CWC22 in the spliceosome remains to be determined.

Despite great progress in elucidating the later steps of EJC assembly, the details of earlier stages are vague, particularly with respect to the MAGOH-RBM8A heterodimer. In good agreement with previous biochemical data, the stable interaction of MAGOH-RBM8A with the C-complex involves the formation of a trimeric

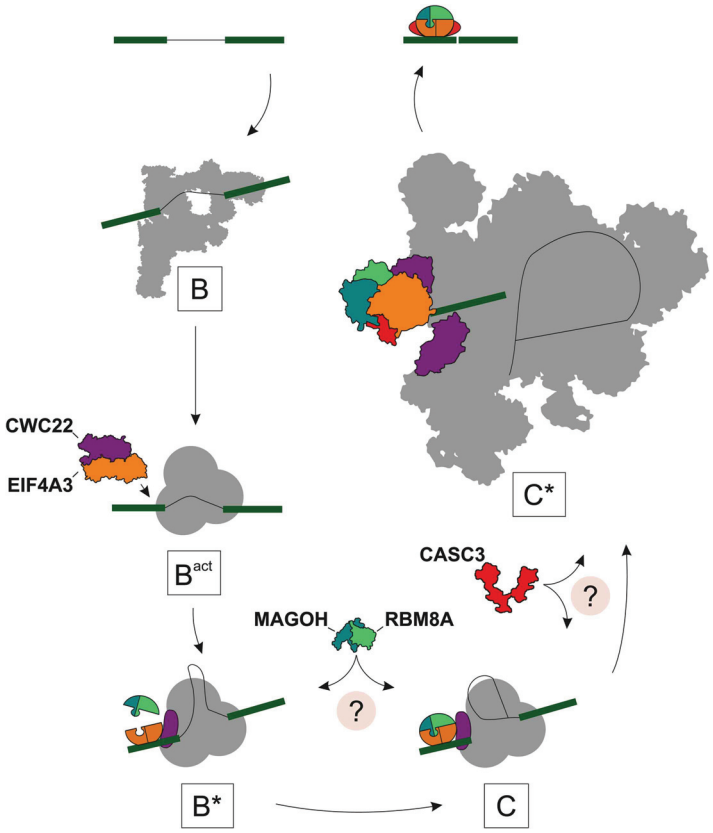


Figure 1. Coupling of splicing and EJC assembly.

The structure of the B complex is schematically depicted based on the PDB file 5O9Z. EIF4A3 (orange) is recruited to the spliceosome by CWC22 (purple), which joins the activated B complex (B^{act}). The CWC22-EIF4A3 structure (PDB 4C9B) is schematically shown. The RBM8A-MAGOH heterodimer joins the EJC in spliceosomal complex B* or C, whereas CASC3 probably associates shortly afterwards. In the structure of the C* spliceosome (PDB 5XJC), the positioning of all EJC components is known. After splicing, the EJC remains bound upstream of the exon-exon junction.

pre-EJC comprising MAGOH-RBM8A and EIF4A3 [30]. It seems plausible that MAGOH-RBM8A also contacts earlier spliceosomal complexes, but the molecular details are not known. To gain further insight into this aspect of EJC assembly, structural details about earlier spliceosomal complexes, such as B* and C, would be needed. These would also aid in establishing the role of other spliceosomal proteins. For example, the intron-binding helicase IBP160 (Aquarius) has been reported to function in EJC assembly [31], but is located on the other side of the EJC in the C* spliceosome.

Table 1 Overview of the protein domains mentioned.

Protein	Domain	Size	Function	Also found in	References
CWC22	MA3	~120 aa	Protein–protein interaction, regulation of DEAD-box proteins	eIF4G, NOM1	[23,28,29]
	MIF4G	~200 aa		eIF4G, NOM1	
EIF4A3	RecA1	~170 aa	ATP binding and hydrolysis, RNA binding	RNA helicases	[13]
	RecA2	~160 aa			
CASC3	SELOR	~150 aa	Protein–protein interaction	–	[16]

Biochemical Society Transactions (2018) **46** 153–161
https://doi.org/10.1042/BST20170059

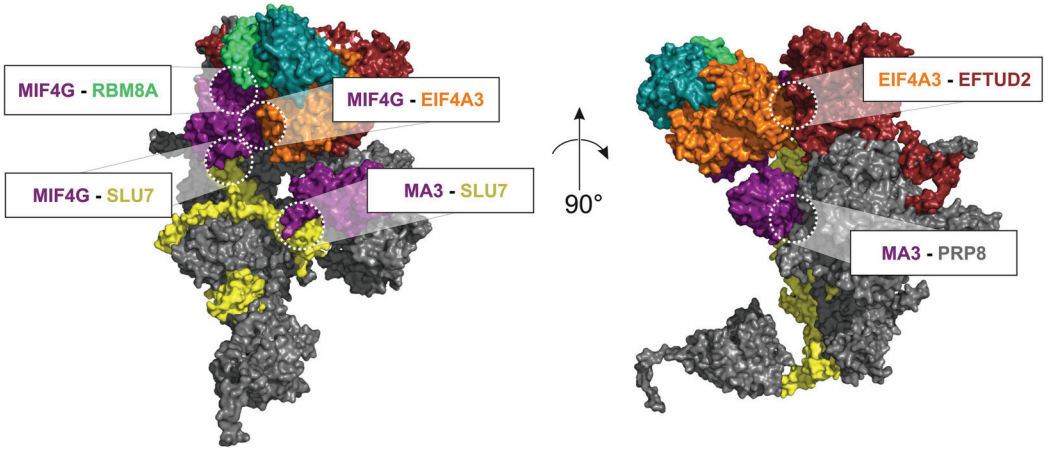


Figure 2. Interactions between EJC components and spliceosomal proteins.

Left: View from the front; Right: View rotated 90°C clockwise about a vertical axis. The interactions of CWC22's domains MIF4G and MA3 (purple) are depicted, as well as the interaction between EIF4A3 (orange) and EFTUD2 (dark red). The other proteins shown are RBM8A (green), MAGOH (blue), SLU7 (yellow) and PRP8 (grey).

Linking the core of the EJC to the functional periphery

The EJC core has the remarkable property of remaining stably associated with the mRNA at the same position, at which it was originally assembled. In order to regulate different steps of gene expression, the EJC associates with peripheral EJC proteins, for which it acts as universal RNA-binding adapter (Figure 3A) [32,33]. The best-understood interactions of peripheral EJC proteins with the EJC core involve the so-called EJC-binding motif (EBM) [34]. This sequence motif was first described in the human NMD factor UPF3B [35] and was later shown to be present in additional EJC interacting proteins, including the human proteins SMG6, TDRD3, GIDRP88, R3HCC1, C2orf68 and RRP12(Figure 3B) [34]. Notably, in most of these proteins, the EBM is located N- or C-terminally, which could facilitate its association with the EJC. On the EJC, the three proteins EIF4A3, RBM8A and MAGOH form a contiguous interface, to which the linear EBM binds (Figure 3C) [36]. The functional consequences of binding of EBM-containing proteins to the EJC were studied only for UPF3B and SMG6, which link the EJC to the NMD-machinery. While UPF3B is believed to attract the NMD factors UPF2 and UPF1 to aberrant mRNAs [37,38], SMG6 is an endonuclease that cleaves mRNA in the vicinity of premature translation termination codons [39,40]. However, how the interaction of two potentially competing NMD factors (UPF3B and SMG6) to the same binding site is regulated remains to be determined.

In addition to its role in NMD, the EJC has been shown to support the export of spliced mRNAs from the nucleus to the cytoplasm [38]. This function is dependent on the interaction of dedicated export factors, specifically UAP56 (DDX39B) and ALYREF, with the EJC [38]. While little is known about the binding of UAP56 to the EJC, the interaction of ALYREF and the EJC involves a short linear motif in ALYREF (WxHD) that was previously identified to mediate binding of CASC3 to EIF4A3 (Figure 3A) [41]. Therefore, binding of ALYREF to spliced mRNAs in the nucleus requires EIF4A3 as part of the EJC as well as the cap-binding complex (CBP80/CBP20), which provides an additional binding site and stimulates export [42].

Two additional complexes associate with the EJC, consisting of the proteins RNPS1, SAP18 and either ACIN1 (Acinus, referred to as ASAP complex) or PNN (Pinin, termed PSAP complex) [32,43]. Both complexes are composed of known regulators of splicing and may therefore influence the excision of specific introns in an EJC-dependent manner. How these two complexes are recruited to the EJC and if their interaction requires additional proteins are currently not known (Figure 3A).

If all these peripheral factors were able to interact with the EJC simultaneously, they may unintendedly compete with each other. Hence, the dynamic association and dissociation of peripheral EJC proteins is required in order to regulate all EJC-dependent processes. How the composition of an EJC at a given time point is determined remains to be elucidated, but the local concentrations of peripheral components probably play a decisive role. In addition, it is conceivable that specific processes exist to detach or replace proteins that are no longer needed.

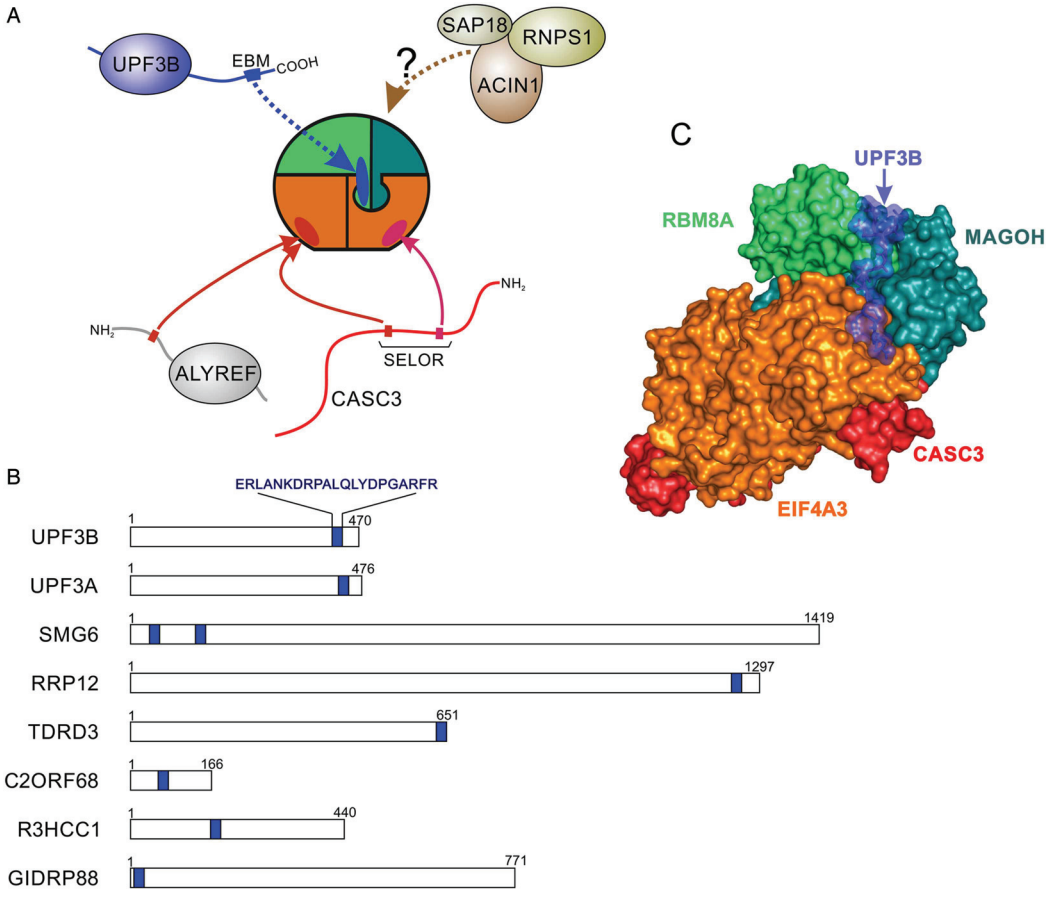


Figure 3. Interactions of peripheral EJC proteins with the EJC core.

(**A**) Model of peripheral EJC proteins binding to the EJC core. Three different types of interactions are shown as examples. UPF3B and other EBM-containing proteins interact with the EJC core via a composite-binding site formed by EIF4A3, RBM8A and MAGOH. A short linear motif in ALYREF (WxHD), also present in the EJC protein CASC3, mediates binding to EIF4A3. Molecular specifications of the interaction of ASAP (ACIN1, RNPS1, SAP18) and PSAP (Pinin, RNPS1, SAP18; not shown) complexes with the EJC core are unknown. For more details, see main text. (**B**) Schematic representation of human EBM-containing proteins reported by Kashima et al. [37]. EBM motifs are displayed as blue boxes. (**C**) Crystal structure showing details of the interaction of the UPF3B EBM peptide (blue) with the core of the EJC. The image was created with PyMOL using the PDB file 2XB2.

EJC disassembly

The amount of EJC protein components in the cell is limited in comparison to the high number of exon-exon junctions present in transcripts [44–46]. Therefore, EJC subunits must be efficiently recycled to provide a sufficient supply of EJC components for the assembly of EJCs on all spliced nascent transcripts. Because EJCs aid in defining premature termination codons in mRNA surveillance, they should remain assembled on exported mRNA in the cytoplasm until the first round of translation. Indeed, RBM8A has been shown to be associated with spliced mRNA in the cytoplasm until it is removed during translation [2]. It is conceivable that EJCs, like many other mRNA-bound proteins, are removed by the processivity of the ribosome, which possesses mRNA helicase activity [47].

The identification of PYM (Partner of Y14/RBM8A-MAGOH) as an EJC disassembly factor provided a new perspective on the mechanism of EJC disassembly [46]. It could be shown that PYM is able to dissociate assembled EJCs from spliced mRNAs *in vitro* and *in vivo*. Depletion of PYM increases the amount of EJCs associated with mRNA, while impairing recycling of EJC components.

Biochemical Society Transactions (2018) **46** 153–161
https://doi.org/10.1042/BST20170059

The molecular mechanism of PYM-mediated EJC disassembly involves transient interaction of PYM with EJC components. PYM was first identified in *Drosophila* S2 cells as part of a trimeric complex with RBM8A and MAGOH [48]. The crystal structure of the RBM8A-MAGOH-PYM complex from *Drosophila* revealed that the N-terminus of PYM interacts with a region formed by both RBM8A and MAGOH [49]. However, the RBM8A-MAGOH heterodimer is not able to bind PYM while being present in the EJC, since residues involved in PYM binding normally interact with EIF4A3 [12]. Interestingly, PYM cannot interact with RBM8A-MAGOH during EJC assembly and only mature EJCs are susceptible to PYM-binding. The ability of PYM to discriminate mature EJCs must be regulated to protect nascent EJCs from disassembly. This is aided by the fact that PYM is mainly located in the cytoplasm [49]. Furthermore, human PYM has been shown to associate with the 40S ribosomal subunit [50] which could serve to couple EJC disassembly to translation.

The importance of EJC assembly by PYM *in vivo* has been addressed in *Drosophila*, where the EJC is required for oskar mRNA localization during oogenesis [51]. Using transgenic approaches, it was shown that increased amounts of *Drosophila* PYM (DmPYM) lead to the dissociation of EJCs from oskar mRNA. However, unlike human PYM, the activity of DmPYM is not coupled to translation and it does not interact with the ribosome. Rather, the regulation of DmPYM seems mediated via its C-terminus [52]. Notably, DmPYM is dispensable in wild-type *Drosophila*, but becomes essential in flies that express reduced levels of y14 (RBM8A in humans) or mago nashi (MAGOH in humans). Therefore, one of the main functions of PYM *in vivo* could be to maintain the balance between EJC components and assembled EJCs.

Conclusion and future directions

The recently reported cryo-EM structures have not only improved the mechanistic understanding of EJC positioning by the spliceosome but also revealed the previously unknown interactions of CWC22 with PRP8, EFTUD2 and SLU7. The question remains whether these factors aid in positioning CWC22 in the spliceosome and if additional factors are involved in the recruitment of CWC22. In addition, it will be important to follow the route of CWC22 after completion of the splicing reaction, since it has to re-associate with EIF4A3 before the next round of splicing. First of all, it is not clear, how exactly CWC22 dissociates from the spliceosome. Studies in yeast revealed that when the post-catalytic spliceosome is dissociated into mRNA and intron-lariat spliceosome by Prp22p (DHX8 in humans), Cwc22p does not have a high affinity for the intron-lariat spliceosome [53]. This points to the possibility that Prp22p has a function in Cwc22p recycling, which is supported by the observation that EJC assembly is disrupted by an ATPase-deficient mutant of DHX8/HRH1 in humans [54]. The position of DHX8 close to CWC22 in the human C* spliceosome also hints at a potential involvement. The exact connection between human DHX8 and CWC22 in recycling and thus EJC assembly remains to be determined.

Furthermore, the structures of the C* spliceosome have provided information about previously unidentified interactions between the EJC and spliceosomal proteins. This includes the interaction between EFTUD2 and EIF4A3 and a novel interaction between CWC22 and RBM8A. To what extent these interactions contribute to the assembly or stability of the EJC still needs to be investigated. Nonetheless, these interactions indicate that EJC assembly occurs via several intermediate stages. In order to characterize these intermediates, however, we will need to understand how and when exactly the EJC components MAGOH, RBM8A and CASC3 are recruited to the spliceosome. While structures of the human B and C* spliceosome are available, structures from the B^{act} and C complex would allow the disentanglement of the multiple steps involved in EJC assembly.

The disassembly of the EJC from mRNA also warrants further investigations. So far, data on the *in vivo* function of PYM have been gathered only in flies. Whether PYM plays a similar role in mammals could be investigated using transgenic or knockout mice. Since the EJC has diverse functions in mammalian cells, PYM may be more important here than in flies. Also on the molecular level, there is a lack of clarity. For example, the connection between EJC disassembly and translation has been discussed controversially. Indeed, the processivity of the ribosome (or the small subunit) should be sufficient to dismantle EJCs during translation (or scanning). The main reason for linking the activity of PYM to translation could be to prevent premature EJC disassembly. It will be an important subject for future studies to investigate which mechanisms exist to maintain the pool of available EJC components and how EJC disassembly is coupled with the reimport of EJC proteins into the nucleus.

Abbreviations

cryo-EM, cryo-electron microscopy; DmPYM, *Drosophila* PYM; EBM, EJC-binding motif; EJC, exon junction complex; MIF4G, middle of initiation factor 4G; NMD, nonsense-mediated mRNA decay; nt, nucleotides; PYM, Partner of Y14-MAGOH

Author Contribution

All authors wrote and edited the manuscript.

Funding

This research was funded by a Heisenberg-Fellowship [GE2014/5-1] and a grant from the Deutsche Forschungsgemeinschaft [GE2014/6-1] to N.H.G.

Acknowledgements

We apologize to all authors whose work was inadvertently overlooked. We thank Volker Böhm for his critical reading of the manuscript and many helpful discussions.

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

References

1 Gehring, N.H., Wahle, E. and Fischer, U. (2017) Deciphering the mRNP code: RNA-Bound determinants of post-transcriptional gene regulation. *Trends Biochem. Sci.* **42**, 369–382 https://doi.org/10.1016/j.tibs.2017.02.004
2 Dostie, J. and Dreyfuss, G. (2002) Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr Biol.* **12**, 1060–1067 https://doi.org/10.1016/S0960-9822(02)00902-8
3 Lejeune, F., Ishigaki, Y., Li, X. and Maquat, L.E. (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J.* **21**, 3536–3545 https://doi.org/10.1093/emboj/cdf345
4 Boehm, V. and Gehring, N.H. (2016) Exon junction complexes: supervising the gene expression assembly line. *Trends Genet.* **32**, 724–735 https://doi.org/10.1016/j.tig.2016.09.003
5 Le Hir, H., Saulière, J. and Wang, Z. (2016) The exon junction complex as a node of post-transcriptional networks. *Nat. Rev. Mol. Cell Biol.* **17**, 41–54 https://doi.org/10.1038/nrm.2015.7
6 Gallego-Paez, L.M., Bordone, M.C., Leote, A.C., Saraiva-Agostinho, N., Ascensao-Ferreira, M. and Barbosa-Morais, N.L. (2017) Alternative splicing: the pledge, the turn, and the prestige: the key role of alternative splicing in human biological systems. *Hum. Genet.* **136**, 1015–1042 https://doi.org/10.1007/s00439-017-1790-y
7 Wahl, M.C., Will, C.L. and Lührmann, R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 https://doi.org/10.1016/j.cell.2009.02.009
8 Will, C.L. and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 https://doi.org/10.1101/cshperspect.a003707
9 Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J. and Shi, Y. (2017) An atomic structure of the human spliceosome. *Cell* **169**, 918–929.e14 https://doi.org/10.1016/j.cell.2017.04.033
10 Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K. et al. (2017) Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**, 318–323 https://doi.org/10.1038/nature21079
11 Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L. et al. (2017) Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. *Cell* **170**, 701–713.e11 https://doi.org/10.1016/j.cell.2017.07.011
12 Bono, F., Ebert, J., Lorentzen, E. and Conti, E. (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* **126**, 713–725 https://doi.org/10.1016/j.cell.2006.08.006
13 Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L. et al. (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* **313**, 1968–1972 https://doi.org/10.1126/science.1131981
14 Ballut, L., Marchadier, B., Baguet, A., Tomasetto, C., Seraphin, B. and Le Hir, H. (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat. Struct. Mol. Biol.* **12**, 861–869 https://doi.org/10.1038/nsmb990
15 Singh, K.K., Wachsmuth, L., Kulozik, A.E. and Gehring, N.H. (2013) Two mammalian MAGOH genes contribute to exon junction complex composition and nonsense-mediated decay. *RNA Biol.* **10**, 1291–1298 https://doi.org/10.4161/rna.25827
16 Degot, S., Le Hir, H., Alpy, F., Kedinger, V., Stoll, I., Wendling, C. et al. (2004) Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module. *J. Biol. Chem.* **279**, 33702–33715 https://doi.org/10.1074/jbc.M402754200
17 Baguet, A., Degot, S., Cougot, N., Bertrand, E., Chenard, M.-P., Wendling, C. et al. (2007) The exon-junction-complex-component metastatic lymph node 51 functions in stress-granule assembly. *J. Cell Sci.* **120**, 2774–2784 https://doi.org/10.1242/jcs.009225
18 Mao, H., Brown, H.E. and Silver, D.L. (2017) Mouse models of *Casc3* reveal developmental functions distinct from other components of the exon junction complex. *RNA* **23**, 23–31 https://doi.org/10.1261/rna.058826.116
19 Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrançois, M., Chagnon, P. et al. (2010) The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in *Drosophila*. *Cell* **143**, 251–262 https://doi.org/10.1016/j.cell.2010.09.014

Biochemical Society Transactions (2018) **46** 153–161
https://doi.org/10.1042/BST20170059

20 Roignant, J.-Y. and Treisman, J.E. (2010) Exon junction complex subunits are required to splice *Drosophila* MAP kinase, a large heterochromatic gene. *Cell* **143**, 238–250 https://doi.org/10.1016/j.cell.2010.09.036

21 Nielsen, K.H., Chamieh, H., Andersen, C.B.F., Fredslund, F., Hamborg, K., Le Hir, H. et al. (2009) Mechanism of ATP turnover inhibition in the EJC. *RNA* **15**, 67–75 https://doi.org/10.1261/rna.1283109

22 Le Hir, H., Izaurralde, E., Maquat, L.E. and Moore, M.J. (2000) The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* **19**, 6860–6869 https://doi.org/10.1093/emboj/19.24.6860

23 Steckelberg, A.-L., Boehm, V., Gromadzka, A.M. and Gehring, N.H. (2012) CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep.* **2**, 454–461 https://doi.org/10.1016/j.celrep.2012.08.017

24 Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M. et al. (2012) Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat. Struct. Mol. Biol.* **19**, 983–990 https://doi.org/10.1038/nsmb.2380

25 Alexandrov, A., Colognori, D., Shu, M.-D. and Steitz, J.A. (2012) Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc. Natl Acad. Sci. U.S.A.* **109**, 21313–21318 https://doi.org/10.1073/pnas.1219725110

26 Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L. et al. (2011) Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* **31**, 2667–2682 https://doi.org/10.1128/MCB.05266-11

27 Steckelberg, A.-L., Altmueller, J., Dieterich, C. and Gehring, N.H. (2015) CWC22-dependent pre-mRNA splicing and eIF4A3 binding enables global deposition of exon junction complexes. *Nucleic Acids Res.* **43**, 4687–4700 https://doi.org/10.1093/nar/gkv320

28 Marintchev, A., Edmonds, K.A., Marintcheva, B., Hendrickson, E., Oberer, M., Suzuki, C. et al. (2009) Topology and regulation of the human eIF4A/4G/4H helicase complex in translation initiation. *Cell* **136**, 447–460 https://doi.org/10.1016/j.cell.2009.01.014

29 Alexandrov, A., Colognori, D. and Steitz, J.A. (2011) Human eIF4AIII interacts with an eIF4G-like partner, NOM1, revealing an evolutionarily conserved function outside the exon junction complex. *Genes Dev.* **25**, 1078–1090 https://doi.org/10.1101/gad.2045411

30 Gehring, N.H., Lamprinaki, S., Hentze, M.W. and Kulozik, A.E. (2009) The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol.* **7**, e1000120 https://doi.org/10.1371/journal.pbio.1000120

31 Ideue, T., Sasaki, Y.T.F., Hagiwara, M. and Hirose, T. (2007) Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev.* **21**, 1993–1998 https://doi.org/10.1101/gad.1557907

32 Tange, T.O., Shibuya, T., Jurica, M.S. and Moore, M.J. (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* **11**, 1869–1883 https://doi.org/10.1261/rna.2155905

33 Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z. et al. (2012) The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* **151**, 750–764 https://doi.org/10.1016/j.cell.2012.10.007

34 Kashima, I., Jonas, S., Jayachandran, U., Buchwald, G., Conti, E., Lupas, A.N. et al. (2010) SMG6 interacts with the exon junction complex via two conserved EJC-binding motifs (EBMs) required for nonsense-mediated mRNA decay. *Genes Dev.* **24**, 2440–2450 https://doi.org/10.1101/gad.604610

35 Gehring, N.H., Neu-Yilik, G., Schell, T., Hentze, M.W. and Kulozik, A.E. (2003) Y14 and hUpf3b form an NMD-activating complex. *Mol. Cell* **11**, 939–949 https://doi.org/10.1016/S1097-2765(03)00142-4

36 Buchwald, G., Ebert, J., Basquin, C., Sauliere, J., Jayachandran, U., Bono, F. et al. (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc. Natl Acad. Sci. U.S.A.* **107**, 10050–10055 https://doi.org/10.1073/pnas.1000993107

37 Chamieh, H., Ballut, L., Bonneau, F. and Le Hir, H. (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.* **15**, 85–93 https://doi.org/10.1038/nsmb1330

38 Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M.J. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**, 4987–4997 https://doi.org/10.1093/emboj/20.17.4987

39 Boehm, V., Haberman, N., Ottens, F., Ule, J. and Gehring, N.H. (2014) 3′ UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep.* **9**, 555–568 https://doi.org/10.1016/j.celrep.2014.09.012

40 Eberle, A.B., Lykke-Andersen, S., Mühlemann, O. and Jensen, T.H. (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.* **16**, 49–55 https://doi.org/10.1038/nsmb.1530

41 Gromadzka, A.M., Steckelberg, A.-L., Singh, K.K., Hofmann, K. and Gehring, N.H. (2016) A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucleic Acids Res.* **44**, 2348–2361 https://doi.org/10.1093/nar/gkw009

42 Cheng, H., Dufu, K., Lee, C.-S., Hsu, J.L., Dias, A. and Reed, R. (2006) Human mRNA export machinery recruited to the 5′ end of mRNA. *Cell* **127**, 1389–1400 https://doi.org/10.1016/j.cell.2006.10.044

43 Murachelli, A.G., Ebert, J., Basquin, C., Le Hir, H. and Conti, E. (2012) The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. *Nat. Struct. Mol. Biol.* **19**, 378–386 https://doi.org/10.1038/nsmb.2242

44 Deutsch, M. and Long, M. (1999) Intron--exon structures of eukaryotic model organisms. *Nucleic Acids Res.* **27**, 3219–3228 https://doi.org/10.1093/nar/27.15.3219

45 Jackson, D.A., Iborra, F.J., Manders, E.M.M. and Cook, P.R. (1998) Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523–1536 https://doi.org/10.1091/mbc.9.6.1523

46 Gehring, N.H., Lamprinaki, S., Kulozik, A.E. and Hentze, M.W. (2009) Disassembly of exon junction complexes by PYM. *Cell* **137**, 536–548 https://doi.org/10.1016/j.cell.2009.02.042

47 Takyar, S., Hickerson, R.P. and Noller, H.F. (2005) mRNA helicase activity of the ribosome. *Cell* **120**, 49–58 https://doi.org/10.1016/j.cell.2004.11.042

48 Forler, D., Köcher, T., Rode, M., Gentzel, M., Izaurralde, E. and Wilm, M. (2003) An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nat. Biotechnol.* **21**, 89–92 https://doi.org/10.1038/nbt773

49 Bono, F., Ebert, J., Unterholzner, L., Güttler, T., Izaurralde, E. and Conti, E. (2004) Molecular insights into the interaction of PYM with the Mago-Y14 core of the exon junction complex. *EMBO Rep.* **5**, 304–310 https://doi.org/10.1038/sj.embor.7400091

50 Diem, M.D., Chan, C.C., Younis, I. and Dreyfuss, G. (2007) PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat. Struct. Mol. Biol.* **14**, 1173–1179 https://doi.org/10.1038/nsmb1321

51 Hachet, O. and Ephrussi, A. (2004) Splicing of oskar RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* **428**, 959–963 https://doi.org/10.1038/nature02521

52 Ghosh, S., Obrdlik, A., Marchand, V. and Ephrussi, A. (2014) The EJC binding and dissociating activity of PYM is regulated in *Drosophila*. *PLoS Genet.* **10**, e1004455 https://doi.org/10.1371/journal.pgen.1004455

53 Fourmann, J.-B., Schmitzova, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.-L. et al. (2013) Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev.* **27**, 413–428 https://doi.org/10.1101/gad.207779.112

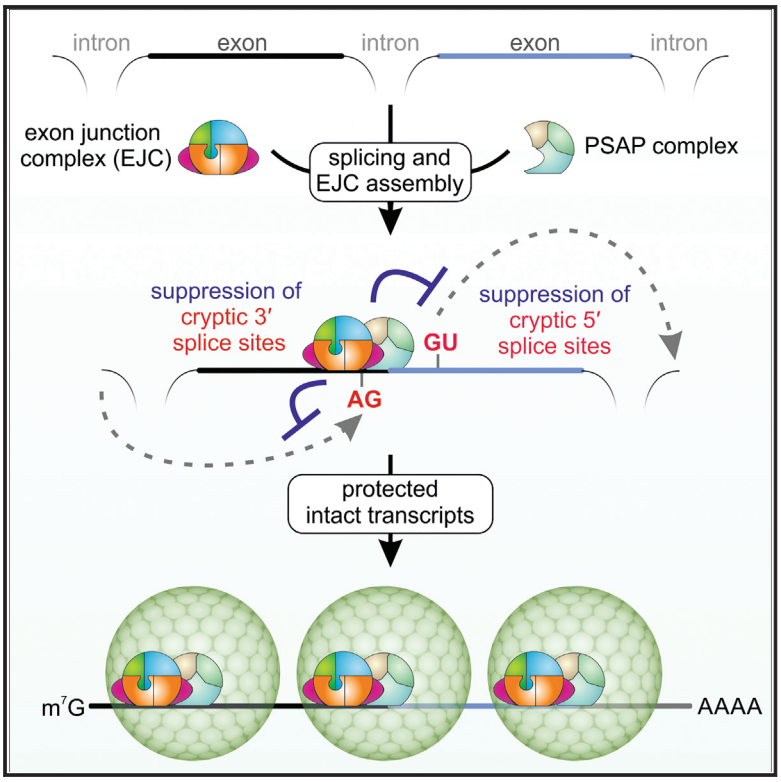
54 Zhang, Z. and Krainer, A.R. (2007) Splicing remodels messenger ribonucleoprotein architecture via eIF4A3-dependent and -independent recruitment of exon junction complex components. *Proc. Natl Acad. Sci. U.S.A.* **104**, 11574–11579 https://doi.org/10.1073/pnas.0704946104

Article

Molecular Cell

Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity

Graphical Abstract



Highlights

- Reduced EJC abundance results in aberrant splicing using cryptic splice sites
- PSAP/EJC-associated RNPS1 suppresses cryptic and reconstituted 5′ splice sites
- RNA-bound EJC core factors directly mask cryptic 3′ splice sites
- RNPS1 and the EJC protect spliced exons by preventing re-splicing

Authors

Volker Boehm, Thiago Britto-Borges, Anna-Lena Steckelberg, ..., Janine Altmüller, Christoph Dieterich, Niels H. Gehring

Correspondence

christoph.dieterich@uni-heidelberg.de (C.D.), ngehring@uni-koeln.de (N.H.G.)

In Brief

The recognition of authentic splice sites is important for the accurate maturation of pre-mRNAs. Boehm et al. demonstrate that the splicing-dependent deposition of exon junction complexes (EJCs) on mRNAs suppresses nearby cryptic splice sites. Thereby, the EJC precludes re-splicing of already spliced regions and maintains the expression of intact transcripts.

Boehm et al., 2018, Molecular Cell 72, 482–495
November 1, 2018
https://doi.org/10.1016/j.molcel.2018.08.030

CellPress

CellPress

Molecular Cell
Article

Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity

Volker Boehm,[1] Thiago Britto-Borges,[2,3] Anna-Lena Steckelberg,[1,4] Kusum K. Singh,[1,5] Jennifer V. Gerbracht,[1] Elif Gueney,[1] Lorea Blazquez,[6,7] Janine Altmüller,[8,9,10] Christoph Dieterich,[2,3,*] and Niels H. Gehring[1,11,*]

[1]Institute for Genetics, University of Cologne, 50674 Cologne, Germany
[2]Section of Bioinformatics and Systems Cardiology, Department of Internal Medicine III and Klaus Tschira Institute for Integrative Computational Cardiology, University of Heidelberg, 69120 Heidelberg, Germany
[3]DZHK (German Centre for Cardiovascular Research), Partner Site Heidelberg/Mannheim, 69120 Heidelberg, Germany
[4]Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, CO 80045, USA
[5]Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, 781039-Guwahati, Assam, India
[6]Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
[7]The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
[8]Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, Germany
[9]Institute of Human Genetics, University of Cologne, 50931 Cologne, Germany
[10]Center for Molecular Medicine Cologne, University of Cologne, 50937 Cologne, Germany
[11]Lead Contact
*Correspondence: christoph.dieterich@uni-heidelberg.de (C.D.), ngehring@uni-koeln.de (N.H.G.)
https://doi.org/10.1016/j.molcel.2018.08.030

SUMMARY

Productive splicing of human precursor messenger RNAs (pre-mRNAs) requires the correct selection of authentic splice sites (SS) from the large pool of potential SS. Although SS consensus sequence and splicing regulatory proteins are known to influence SS usage, the mechanisms ensuring the effective suppression of cryptic SS are insufficiently explored. Here, we find that many aberrant exonic SS are efficiently silenced by the exon junction complex (EJC), a multi-protein complex that is deposited on spliced mRNA near the exon-exon junction. Upon depletion of EJC proteins, cryptic SS are de-repressed, leading to the mis-splicing of a broad set of mRNAs. Mechanistically, the EJC-mediated recruitment of the splicing regulator RNPS1 inhibits cryptic 5′SS usage, while the deposition of the EJC core directly masks reconstituted 3′SS, thereby precluding transcript disintegration. Thus, the EJC protects the transcriptome of mammalian cells from inadvertent loss of exonic sequences and safeguards the expression of intact, full-length mRNAs.

INTRODUCTION

The majority of eukaryotic pre-mRNAs undergo alternative splicing and produce an assorted set of mRNAs (Lee and Rio, 2015). Splicing of pre-mRNAs not only increases the coding potential of the genome but is also a key regulatory step in gene expression (Kornblihtt et al., 2013; Nilsen and Graveley, 2010). In order to execute the two consecutive steps of splicing, the splicing machinery has to correctly identify the splice sites (SS) at the 5′ and 3′ ends of the intron, which have defined consensus sequences (Papasaikas and Valcárcel, 2016). Due to the degenerate nature of the splice consensus sequences, the cell is faced with the challenging task to discriminate between authentic and so-called cryptic SS, which exhibit consensus motifs but are not intended to be used. Therefore, many different mRNA-binding proteins assist the spliceosome in the accurate detection of introns and SS. These splicing regulators commonly bind to specific sequence motifs on the transcript and act as enhancers (e.g., SR proteins) or silencers (e.g., hnRNP proteins) of splicing (Han et al., 2010; Long and Caceres, 2009).

In addition to the removal of intronic sequences, splicing also alters the protein composition of the messenger ribonucleoprotein (mRNP). This phenomenon is documented, in particular, for the exon junction complex (EJC), which is assembled and deposited onto mRNAs during splicing (Boehm and Gehring, 2016; Le Hir et al., 2016). Binding of the EJC to its canonical site 24 nucleotides (nt) upstream of exon-exon junctions does not require a specific RNA-sequence and involves the phosphate backbone of the RNA (Andersen et al., 2006; Bono et al., 2006). The core of the EJC consists of four proteins (EIF4A3, MAGOH, RBM8A, and CASC3) that can be used as an assembly platform for other proteins, the so-called peripheral EJC components (Singh et al., 2012). The deposition of the EJC is initiated by the recruitment of the core factor EIF4A3 to the activated spliceosome by the splicing factor CWC22 (Alexandrov et al., 2012; Barbosa et al., 2012; Steckelberg et al., 2012).

On the cellular level, EJCs represent central mRNP components with diverse functions. Specifically, the EJC serves as the molecular memory of the splicing process and passes on this information to later steps of gene expression (Le Hir et al., 2016; Woodward et al., 2017). In particular, the dynamically

recruited peripheral EJC proteins expand the functional impact of the EJC on gene expression. EJC components have been shown to stimulate mRNA transport, increase translation efficiency, and support mRNA surveillance by nonsense-mediated mRNA decay (NMD) (Boehm and Gehring, 2016; Le Hir et al., 2016; Woodward et al., 2017). With these functions, the EJC helps to ensure that correctly processed and error-free transcripts are preferentially expressed. In addition to these post-splicing processes, the EJC has also been shown to influence the splicing process of selected mRNAs. For instance, correct splicing of *mapk* pre-mRNA and other long intron-containing transcripts in *Drosophila* was found to require EJC components (Ashton-Beaucage et al., 2010; Roignant and Treisman, 2010). Furthermore, depletion of EJC core factors resulted in the retention of a suboptimal intron in the *piwi* transcript (Hayashi et al., 2014; Malone et al., 2014). It has been suggested that in this case splicing of the neighboring introns leads to the deposition of EJCs, which subsequently function as splicing enhancers for the weak intron. In human cells, depletion of EJC core components caused widespread changes in alternative pre-mRNA splicing (Wang et al., 2014). Different types of alternative splicing events were observed, of which cassette exons represented the majority.

In mammals, the known splicing regulators ACIN1, PNN, RNPS1, and SAP18 are peripheral EJC components, which have been shown to co-purify with the EJC core (Singh et al., 2012). Interestingly, RNPS1 and SAP18 form two alternative complexes with either ACIN1 (also known as Acinus) or PNN (also known as Pinin), referred to as apoptosis- and splicing-associated protein (ASAP) or PSAP complex, respectively (Murachelli et al., 2012). Evidence from studies in *Drosophila* suggests that ACIN1 and RNPS1 aid in definition and splicing of neighboring introns and are involved in the EJC-mediated splicing regulation (Hayashi et al., 2014; Malone et al., 2014). Furthermore, RNPS1 is also required for the correct splicing of *mapk* and *AURKB* pre-mRNA in *Drosophila* and human cells, respectively (Ashton-Beaucage et al., 2010; Fukumura et al., 2018; Roignant and Treisman, 2010). Finally, recent transcriptome-wide studies identified alternative splicing changes in ACIN1- and PNN-depleted cells, suggesting that ASAP and PSAP complexes can regulate certain EJC-dependent and -independent splice events (Rodor et al., 2016; Wang et al., 2018). Despite the identification of various splicing alterations upon EJC depletion in mammalian cells, the molecular mechanism underlying the EJC-mediated splicing regulation is mostly uncharacterized. Hence, it is of fundamental importance to dissect the mechanistic role and functional dependency on core and peripheral EJC components for splicing regulation.

Here, we investigate the mechanism of splicing modulation by the EJC. We show that depletion of the EJC-associated splice factor RNPS1 caused widespread changes in splicing and led to the usage of cryptic and reconstituted 5′SS, which are efficiently repressed in the presence of RNPS1. Moreover, we identified an EJC-dependent, but RNPS1-independent, mechanism that prevents splicing involving the usage of cryptic and reconstituted 3′SS. Taken together, the EJC, in cooperation with RNPS1, prevents the recognition of irregular SS within many transcripts and thus the formation of incorrect mRNAs.

RESULTS

Given the lack of insight into the mechanism of splicing regulation by the EJC, we set out to investigate the function of the EJC-associated splice factor RNPS1, a common component of ASAP and PSAP complexes (Mayeda et al., 1999; Murachelli et al., 2012; Sakashita et al., 2004) (Figures 1A and S1A). First, we tested whether RNPS1 is required for the correct splicing of the *MRPL3* transcript, which shows robust exon 4 skipping upon knockdown of the EJC core protein EIF4A3 (Wang et al., 2014). Interestingly, siRNA-mediated depletion of RNPS1 quantitatively recapitulated the 60- to 70-fold increase of EJC-dependent skipping of *MRPL3* exon 4 (Figures 1B and 1C), suggesting a functional link between EJC- and RNPS1-dependent splicing regulation.

RNPS1 Recruitment to Spliced mRNPs via the RRM Domain Is Required for Splicing Regulation

Recent *in vivo* crosslinking and immunoprecipitation (iCLIP) experiments showed that RNPS1 displays an RNA binding pattern similar to the EJC core protein EIF4A3 (Hauer et al., 2016). Hence, we assumed that RNPS1 is positioned on spliced mRNPs via an interaction with RNA-bound EJCs. To better understand the interaction between RNPS1 and the EJC, we studied ASAP/PSAP complex formation and EIF4A3 binding using co-immunoprecipitation experiments. FLAG-tagged RNPS1 co-precipitated the ASAP/PSAP complex component SAP18 as well as the core EJC protein EIF4A3 (Figure S1B). Mutating a surface exposed patch on RNPS1 (termed 176; Figure S1C) that prevents ASAP/PSAP complex formation and co-precipitation of SAP18 also abolished the interaction with EIF4A3 (Figure S1B), suggesting that RNPS1 interacts with the EJC as part of the fully assembled ASAP or PSAP complex. We further investigated the recruitment of RNPS1 to spliced transcripts using immunoprecipitation of *in vitro* spliced mRNPs (Steckelberg and Gehring, 2014). While the C-terminal RS/P domain of RNPS1 conferred unspecific binding to spliced and unspliced RNA (Figures S1D–S1F), the isolated RRM domain of RNPS1, which is sufficient to form a minimal ASAP/PSAP complex (Murachelli et al., 2012), co-immunoprecipitated exclusively spliced transcripts. Interestingly, RNPS1 (176), which is deficient in ASAP/PSAP complex formation and EIF4A3 binding (Figure 1D), also failed to co-precipitate spliced mRNA (Figures 1E and S1G), indicating that ASAP/PSAP complex formation and splicing-dependent mRNP interaction are functionally linked. Of note, *MRPL3* exon skipping observed upon knockdown of RNPS1 could be rescued by expression of wild-type RNPS1, but not RNPS1 (176) (Figure 1F). We conclude that RNPS1 requires the RRM-mediated formation of an ASAP/PSAP complex and the interaction with the EJC for its specific association with spliced mRNPs to modulate splicing.

RNPS1 Depletion Causes Transcriptome-wide Loss of Exonic Sequences

Having established a molecular link between RNPS1 and the EJC, we next examined the global role of RNPS1 in splicing regulation. To this end, we sequenced RNA from control- and RNPS1-depleted cells as well as RNPS1 knockdown cells that

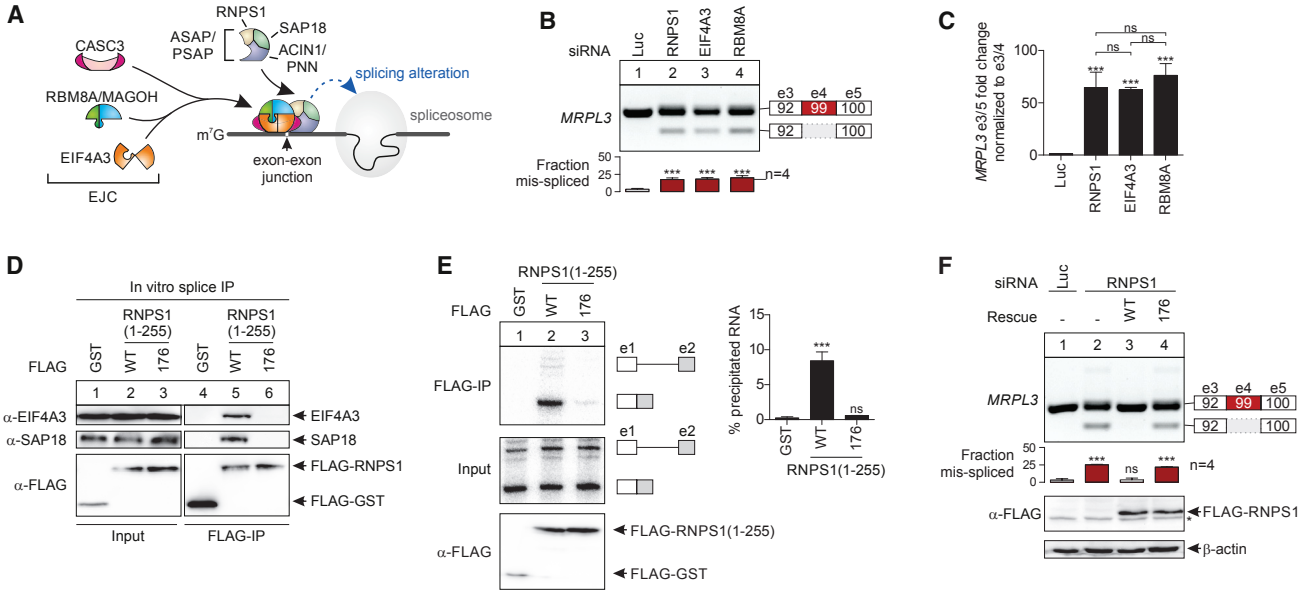


Figure 1. RNPS1 Regulates Splice Site Selection via an Interaction with the Exon Junction Complex

(A) Schematic overview of splicing modulation by exon junction complex (EJC) and ASAP/PSAP-complex formation.

(B) RT-PCR analysis of *MRPL3* exon 4 skipping with RNA from HeLa cells transfected with the indicated siRNA. *MRPL3* exon architecture is depicted schematically; alternatively spliced features are highlighted.

(C) Quantitative RT-PCR analysis of *MRPL3* exon 4 skipping normalized to exon3/4 splicing and compared to the Luc control knockdown. n = 3.

(D) Co-immunoprecipitation of EJC core component EIF4A3 and ASAP/PSAP component SAP18 from *in vitro* splicing experiments with the indicated FLAG-tagged RNPS1 variants. n = 3.

(E) *In vitro* splicing of ^{32}P body-labeled MINX mRNA in the presence of FLAG-RNPS1 variants. n = 3.

(F) RT-PCR analysis of *MRPL3* exon 4 skipping with RNA from stable HeLa cells expressing the indicated rescue proteins, transfected with the indicated siRNA. Western blot analysis of protein expression is shown at the bottom.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

See also Figure S1.

were complemented with RNPS1 wild-type or RNPS1 (176) (Figure 2A). Using the MAJIQ algorithm (Vaquero-Garcia et al., 2016) to identify local splicing variations (LSV), we found that RNPS1 depletion substantially altered the splicing of 318 LSV in 243 genes (Figure 2A; Table S1). The affected genes represented a diverse group, as no specific enrichment was detectable for the gene ontology (GO) terms *molecular function*, *biological process*, or *cellular compartment*.

Classification of the splicing alterations revealed that the predominant group represented exon skipping events, followed by exon inclusion, alternative 5′ or 3′SS usage, and intron retention (Figures 2B and S2A). Remarkably, exon-exon junctions that were barely or never detected under control conditions (herein called "spurious" junctions; see STAR Methods for details) constituted a substantial proportion in all classes of splicing alterations except exon inclusion (Figures 2B and 2C). In more than 30% of the splicing events we observed that RNPS1 depletion leads to the activation of irregular SS, the skipping of constitutive exons, and the formation of unusual transcript variants. Furthermore, none of the RNPS1-dependent splicing alterations were found in a recently published atlas of alternative splicing events in multiple human tissues, cell types, and developmental stages (VAST-DB) (Tapial et al., 2017). We verified the RNA-seq results for selected transcripts with RNPS1-dependent exon skipping, exon inclusion, alternative SS usage, and intron retention by RT-PCR (Figures 2D–2G and S2B–S2I). In all cases, the splicing change caused by RNPS1 knockdown was completely rescued by wild-type RNPS1, but not the RNPS1 (176) mutant, underscoring that RNPS1-dependent splicing events require ASAP/PSAP complex formation and recruitment by the EJC.

RNPS1 Suppresses Cryptic 5′ Splice Sites

The usage of cryptic and irregular SS in RNPS1-depleted cells suggests that the recruitment of RNPS1 by the EJC plays a pivotal role in suppressing these SS under normal conditions. To identify the molecular mechanism of RNPS1-dependent SS suppression, we investigated transcripts with RNPS1-dependent alternative 5′SS, because this class contained a large proportion of spurious junctions with robust fold changes (Figures 2B and 2C). The majority of spurious 5′SS that were upregulated in RNPS1-depleted cells exhibited a good splice consensus sequence (Desmet et al., 2009) and were located close to the 5′ end of an exon (Figure 3A). Therefore, these 5′SS are near canonical EJC-binding sites, on which EJCs can be deposited during splicing of the preceding intron. We hypothesized that RNPS1 bound to an upstream EJC suppresses the usage of nearby 5′SS. To test this model, we took advantage of the *TUFM* transcript, in which usage of a cryptic 5′SS is upregulated

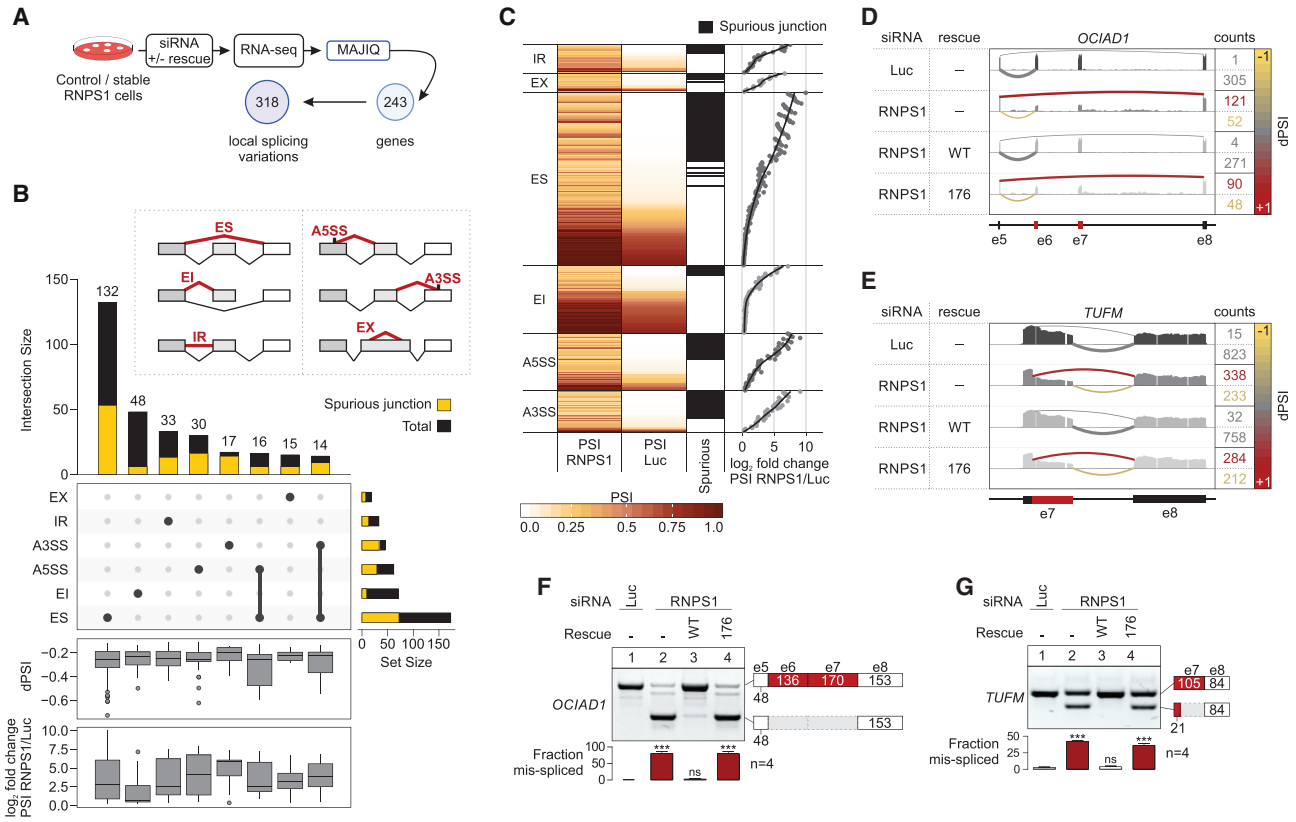


Figure 2. Transcriptome-wide Alternative Splicing Events upon RNPS1 Depletion

(A) Simplified overview of experimental RNA-Seq pipeline.

(B) Classification of selected alternatively spliced junctions upon RNPS1 knockdown as exon skipping (ES), exon inclusion (EI), intron-retention (IR), alternative 5′ or 3′ splice sites (A5SS, A3SS), or exitron (EX) events (see inset). The quantity of spurious junctions for each class is shown in yellow. The change in junction usage (delta percent spliced in, dPSI) and percent spliced in (PSI) fold change for each class are shown as boxplots on the bottom. The full classification plot is shown in Figure S2A.

(C) Heatmaps representing junction usage (percent spliced in; PSI) of alternative splicing events in control (Luc) and RNPS1 knockdown are depicted on the left. The fraction of spurious junctions and the PSI fold change in the same samples are shown on the right. Only junctions with a strong response in RNPS1-depleted cells with a dPSI of <-0.1 were selected.

(D and E) Sashimi-plots of RNA-Seq data of genes with exon skipping (*OCIAD1*, D) and alternative 5′ splice site (SS) usage (*TUFM*, E). Only selected splice junctions are depicted. The thickness and color of the depicted junction represent the junction usage (PSI and dPSI). The counts of reads spanning the indicated junctions are shown.

(F and G) RT- PCR analysis of the alternative splice events shown in (D) and (E) in the *OCIAD1* (F) and the *TUFM* (G) mRNAs, respectively. Total RNA was isolated from stable HeLa cells expressing the indicated rescue proteins and transfected with the indicated siRNA. Quantified results from the indicated biological replicates are shown as mean ± SD and compared to the Luc control knockdown.

See also Figure S2 and Table S1.

upon RNPS1 depletion (Figures 2E and 2G). We constructed a reporter plasmid of the *TUFM* gene, including the RNPS1-dependent 5′SS as well as upstream and downstream introns (Figure 3B). In line with our hypothesis, normal splicing of the reporter mRNA was observed when the intron upstream of the cryptic 5′SS was present (Figure 3C). In contrast, the irregular 5′SS was preferentially used when EJC deposition was prevented by deleting the upstream intron (Figure 3C). These results, together with our data on the position and strength of RNPS1-dependent SS (Figure 3A), suggest that EJC-bound RNPS1 protects a certain region of the downstream exon from the use of irregular 5′SS.

We recognized that our model of RNPS1-dependent 5′SS regulation could also explain exon skipping events, when the skipped exon starts with a potential 5′SS (i.e., contains a GU or GC dinucleotide at the 5′ end as part of a reconstituted SS). In these cases, the same mechanism that regulates cryptic 5′SS could also affect reconstituted SS at the beginning of exons. This idea was supported by the observation that after RNPS1-sensitive exon 4 skipping of the *RER1* mRNA, a single guanosine nucleotide remained (Figures 3D–3F), which could only result from the usage of a reconstituted 5′SS. Indeed, the exon-exon boundary between exon 3 and 4 contains a 5′SS, which can be ligated to the splice acceptor of the downstream intron causing exon 4 skipping (Figure 3G). This re-splicing of a reconstituted exonic 5′SS resembles the process of sequential multi-step splicing of introns termed recursive splicing (Sibley et al., 2016). To test whether exon 4 skipping in the *RER1*

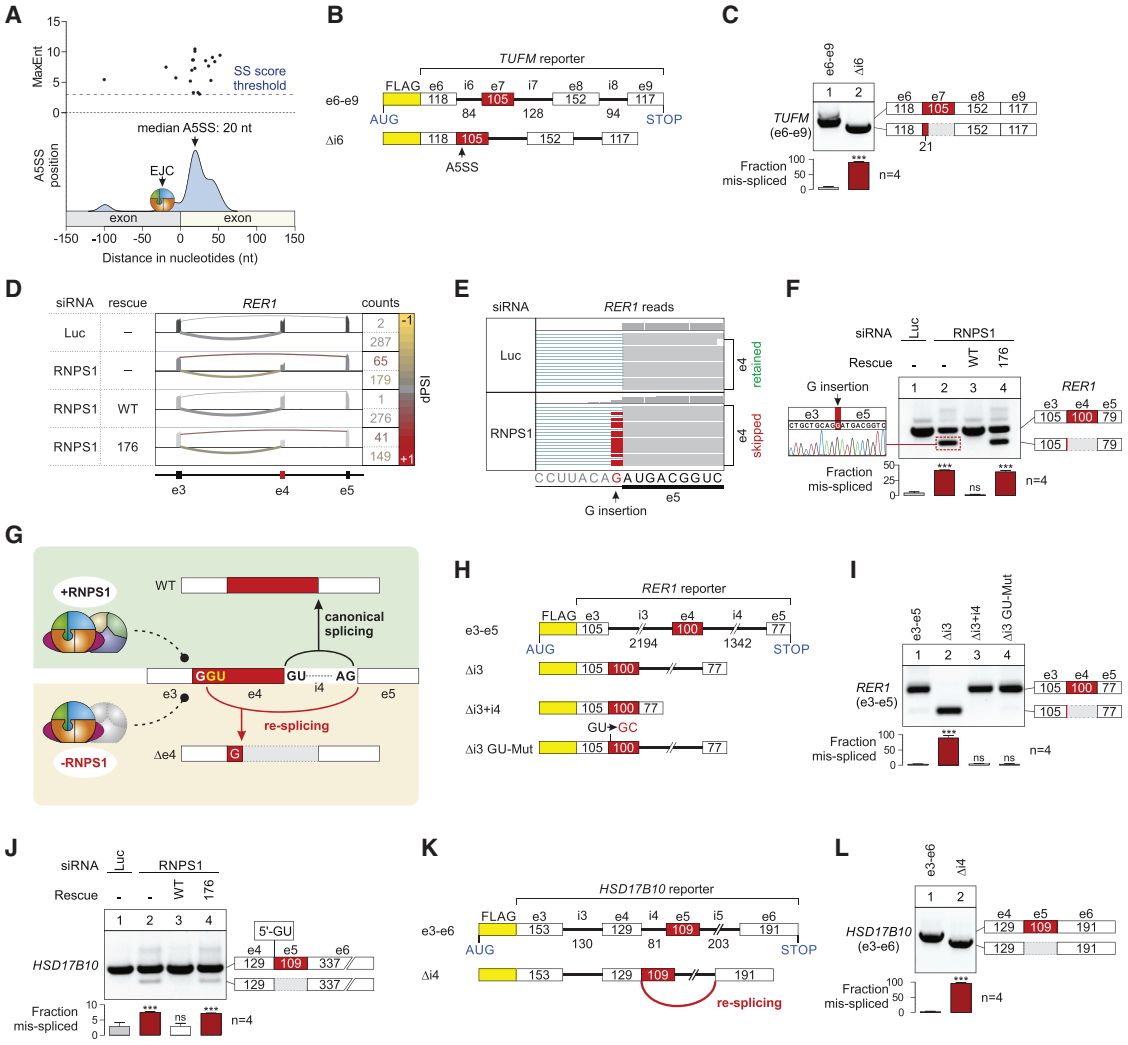


Figure 3. RNPS1 Depletion Leads to Re-splicing of Cryptic and Reconstituted 5′ Splice Sites

(A) Depiction of alternative 5′SS position of spurious junctions relative to exon boundaries as density plot (bottom) and the corresponding 5′ MaxEnt scores (top).
(B and C) *TUFM* minigene reporter constructs (B) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (C).
(D) Sashimi-plots of *RER1* exon 4 skipping from RNA-Seq data.
(E) IGV snapshot of the guanosine insertion for *RER1* reads.
(F) *RER1* RT-PCR of RNA from HeLa cells transfected with the indicated siRNA. Sanger sequencing of the exon 4-skipped *RER1* PCR product is shown with the guanosine retained from exon 4 highlighted.
(G) Scheme of *RER1* exons 3–5 re-splicing in the presence (top) or absence (bottom) of RNPS1.
(H and I) *RER1* minigene reporter constructs (H) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (I).
(J) *HSD17B10* RT-PCR of RNA from HeLa cells transfected with the indicated siRNA. The 5′-terminal GU dinucleotide at the exon 5 of *HSD17B10* is indicated.
(K and L) *HSD17B10* minigene reporter constructs (K) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (L).
All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.
See also Figure S3.

mRNA results from re-splicing, we took advantage of a set of *RER1* reporter plasmids (Figure 3H). When both introns were present in the reporter (e3-e5), exon 4 was constitutively retained, mimicking the constitutive exon inclusion observed *in vivo* (Figures 3H and 3I). A reporter construct lacking the upstream intron 3 (Δi3), which simulates splicing of this intron without EJC-deposition, produced exclusively mRNAs lacking exon 4 (Figures 3H and 3I). Together, these data suggest that intron 3 splicing and concomitant EJC deposition suppresses the use of an adjacent reconstituted SS. When we removed both introns (Δi3+i4) or mutated the GU dinucleotide of the reconstituted 5′SS (Δi3 GU-Mut), exon 4 was retained (Figures 3H and 3I). These findings confirm that in the absence of EJC deposition, the *RER1* transcript undergoes re-splicing using the reconstituted 5′SS and the splice acceptor of intron 4. Strikingly, exon skipping occurred despite the presence of the genuine 5′SS of intron 4

that exhibits a stronger SS consensus score compared to the reconstituted 5′SS (Figure S3A). This observation raises the interesting question of which other determinants affect this SS selection. Apart from the SS consensus sequence, exonic splicing enhancer (ESE) or silencer (ESS) motifs greatly influence the balance of SS usage (Cáceres and Hurst, 2013). Upon inspection of *RER1* exon 4, we suspected that an unfavorable arrangement of ESE and ESS leads to the selection of the reconstituted 5′SS. In support of this hypothesis, removal of several ESSs from the central region of exon 4 and insertion of ESE motifs at the exon's 5′ end gradually restored the usage of the canonical intronic SS in the *RER1* Δi3 reporter (Figures S3A–S3C). These results indicate that RNPS1 and the EJC directly or indirectly counteract exonic splicing silencer motifs, effectively leading to the definition of exons and exon inclusion.

In contrast to *RER1*, where a residual guanosine nucleotide served as a molecular mark of two consecutive splicing events, most exon skipping events observed upon RNPS1 knockdown displayed seamless skipping of one or more exons. We therefore investigated whether these splicing events equally relied on re-splicing or resulted from a direct definition of exon-exon boundaries across multiple introns. To this end, we generated reporter constructs of the *HSD17B10* transcript, for which a moderate increase in exon skipping was observed upon depletion of RNPS1 (Figure 3J). Analogous to *RER1* splicing, near-complete exon skipping occurred in an *HSD17B10* reporter transcript lacking the upstream intron, indicating that this event is indeed a result of re-splicing (Figures 3K and 3L). This remarkable conversion of the splice pattern is surprising, considering that this *HSD17B10* splicing event was near the detection limit of the MAJIQ alternative splicing analysis (fourth last record, Table S1). Importantly, the *HSD17B10* re-splicing event is indistinguishable from regular exon skipping by standard computational analyses, because the first two nucleotides of the skipped exon represent the GU dinucleotide of the 5′SS. This observation suggests that other exon skipping events observed upon RNPS1 knockdown potentially utilize the same re-splicing mechanism. To estimate the probability of seamless re-splicing, we identified the 5′ terminal dinucleotide of the first skipped exon in spurious exon skipping events. More than 75% of the skipped exons exhibited a suitable 5′SS on their 5′ end, suggesting that re-splicing is likely responsible for these exon skipping events (Figure S3D).

The EJC Confers Local Protection from Cryptic Splicing

Based on our observation that compromised EJC deposition results in aberrant splicing, we tested whether site-directed positioning of EJC factors on de-repressed mRNAs can restore normal splicing. To this end, we generated a reporter transcript with MS2 tethering sites to recruit effector proteins to different positions upstream of the reconstituted 5′SS of *HSD17B10* Δi4, which constitutively undergoes re-splicing (Figures 4A and 4B). Tethering of EJC proteins to this mRNA mimics EJC assembly upstream of exon-exon junctions and thus uncouples EJC deposition from pre-mRNA splicing. Indeed, tethering of RNPS1 or the EJC core component RBM8A promoted exon inclusion over a short distance, but gradually lost its effect when the distance between the tethering site and the 5′SS was increased (Figure 4B). These results suggest that MS2-mediated positioning of RNPS1 or EJC components on the mRNA functionally recapitulates splicing-dependent EJC-deposition. Also, our data suggest that the effect of the EJC and RNPS1 is confined to a region surrounding the EJC binding site.

RNPS1 in the PSAP Complex Is the Functional Component for 5′ Splice Site Suppression

To uncover the mechanistic details of cryptic SS suppression by RNPS1, we set out to define the minimal protein domain required for this process. Using rescue assays, we determined that the expression of the isolated RRM domain of RNPS1 considerably rescued RER1 splicing in RNPS1-depleted cells (Figure 4C). This indicates that the isolated RRM domain of RNPS1 retains some functional activity, potentially mediated through the recruitment of other protein factors, such as ASAP/PSAP complex components. To test this, we examined the effects of siRNA-mediated depletion of ASAP/PSAP components. Surprisingly, only the knockdown of SAP18 and PNN, but not ACIN1, resulted in *RER1* exon skipping (Figures 4D and 4E). Furthermore, expression of full-length SAP18 or truncated PNN constructs, which support PSAP formation, restored normal splicing of the *RER1* transcript in SAP18 or PNN-depleted cells, respectively (Figure 4F). Moreover, tethering of SAP18 or PNN to the *HSD17B10* reporter stimulated exon inclusion (Figures S4A and S4B), demonstrating that the PSAP complex is the functional entity involved in EJC-mediated cryptic SS suppression.

To further dissect the hierarchy of EJC- and PSAP-dependent splicing regulation and to identify the active components of the PSAP complex, we performed epistasis experiments using tethering assays in knockdown cells. Tethering of RNPS1 resulted in exon inclusion of the *HSD17B10* reporter mRNA even in EJC- or PSAP-complex depleted cells (Figure 4G). In contrast, tethered PNN and SAP18 were inactive in the absence of RNPS1 (Figure 4H), and tethering of PSAP-incompatible RNPS1 (176) led to similar exon inclusion as tethering of the RNPS1 wild-type protein (Figure 4I). This result supports the hypothesis that RNPS1 is a key effector of EJC- and PSAP-mediated splicing regulation, while EJC and other PSAP components are required primarily for the correct recruitment of RNPS1 to the mRNA.

Based on these findings, we postulate that the PSAP complex, consisting of PNN, SAP18, and RNPS1, contributes to exon definition, a function that suppresses exon skipping and resembles the activity of classical SR proteins (Ibrahim et al., 2005). Supporting this exon definition hypothesis, tethering of SRSF2 or SRSF11, a reported interaction partner of RNPS1 (Sakashita et al., 2004), altered the *HSD17B10* splicing similar to RNPS1 or RBM8A tethering (Figure S4C). However, EJC-dependent exon definition functions independently of these SR proteins, because depletion of SRSF1 and SRSF2 (single or combined) or SRSF11 did not activate RNPS1-dependent cryptic SS (Figures S4D–S4F). Finally, tethered nuclear β-galactosidase (NLS-LacZ, a homotetramer of ~120 kDa proteins) had no effect on exon skipping, indicating that the presence of a large protein complex is not sufficient to prevent exon skipping (Figure S4C). Taken together, these results imply that RNPS1 is an essential effector of SS suppression and is recruited to spliced mRNPs via the formation of PSAP rather than ASAP complexes.

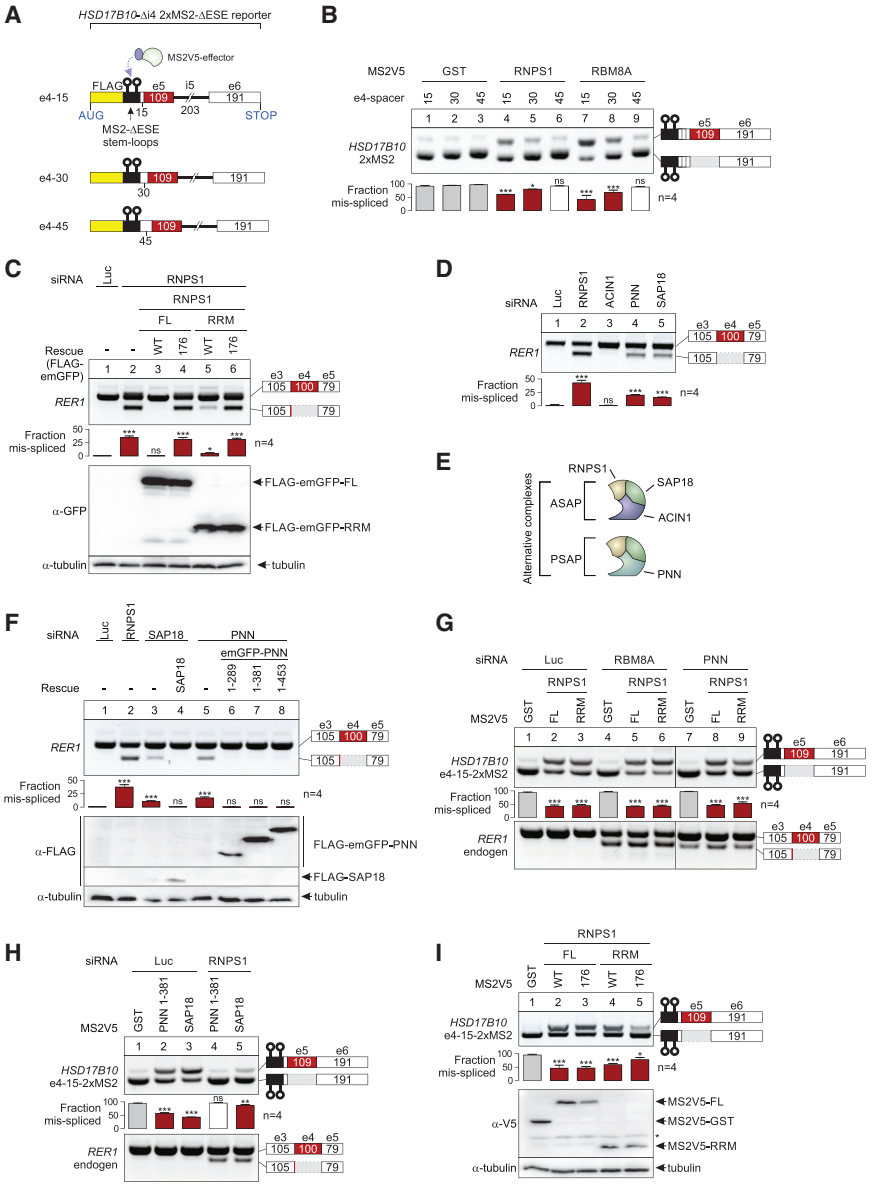


Figure 4. Functional Suppression of Cryptic Splice Sites by the PSAP Complex

(A) Scheme of *HSD17B10* Δi4 tethering reporter. Two ESE-optimized MS2 stem loops were inserted at varying distances upstream of the reconstituted *HSD17B10* cryptic splice site, allowing the direct tethering of MS2V5-tagged proteins.

(B) The indicated MS2V5-tethering proteins and *HSD17B10* Δi4 reporter with varying spacers were expressed transiently in HeLa Tet-Off cells and the splice patterns analyzed by RT-PCR.

(C and F) RT-PCR analysis of *RER1* exon skipping with RNA from stable HeLa cells, expressing the indicated rescue proteins (RNPS1 variants [C] or SAP18 and PNN variants [F]), transfected with the indicated siRNA. Western blot analysis of expressed FLAG- and FLAG-emGFP-tagged proteins is shown. Tubulin served as loading control.

(D) RT-PCR analysis of *RER1* exon skipping with RNA from HeLa cells transfected with the indicated siRNA.

(E) Scheme depicting the components of the ASAP and PSAP complexes.

(G) Dual-inducible stable HeLa cell lines expressing both the *HSD17B10* Δi4 e4-15 reporter and the indicated MS2V5-tagged RNPS1 protein variants were transfected with the indicated siRNA. Splicing of the reporter mRNA and the endogenous *RER1* mRNA was analyzed by RT-PCR.

(H) Dual-inducible stable HeLa cell lines expressing both the *HSD17B10* Δi4 e4-15 reporter and the indicated MS2V5-tagged PNN (1-381) or SAP18 protein were transfected with the indicated siRNA. Splicing of the reporter mRNA and the endogenous *RER1* mRNA was analyzed by RT-PCR.

(I) Dual-inducible stable HeLa cell lines expressing both the *HSD17B10* Δi4 e4-15 reporter and the indicated MS2V5-tagged RNPS1 protein variants were transfected with the indicated siRNA. Splicing of the reporter mRNA was analyzed by RT-PCR.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

See also Figure S4.

EJCs Prevent 3′ Splice Sites Usage in an RNPS1-Independent Manner

Transcripts harboring RNPS1-dependent cryptic or reconstituted SS also showed splicing defects upon depletion of the EJC core components EIF4A3 and RBM8A (Figures 1B and S5A–S5I). Moreover, expression of a dominant-negative splicing factor that is unable to recruit EIF4A3 to the spliceosome (CWC22 [171]) and thus enables pre-mRNA splicing without EJC deposition (Steckelberg et al., 2012) also caused *RER1* exon skipping (Figures S5J and S5K). Hence, a clear link between EJC deposition and splice regulation by RNPS1 exists. However, comparing the identified transcriptome-wide RNPS1-dependent splicing events with previously published RNA-Seq data from EIF4A3-depleted cells (Wang et al., 2014), we noticed that many splicing events induced by EJC knockdown were not found in RNPS1-depleted cells (Figures S5L–S5P; Table S2).

Although it is expected that the EJC knockdown strongly inhibits downstream processes like NMD, leading to the overstabilization of mis-spliced mRNAs, this observation suggested that additional EJC-dependent (but RNPS1-independent) mechanisms act on these mRNAs. The depletion of the EJC core component EIF4A3 mainly caused exon skipping, similar to RNPS1 knockdown (Figures 5A and S6A). However, the second most frequent splicing alteration was the use of alternative 3′SS, an event that was less often observed upon RNPS1 depletion. This observation indicated a mechanistic difference between EJC- and PSAP-dependent splicing regulation and prompted us to investigate this type of splicing dysregulation in more detail. In search of a possible mechanistic explanation for

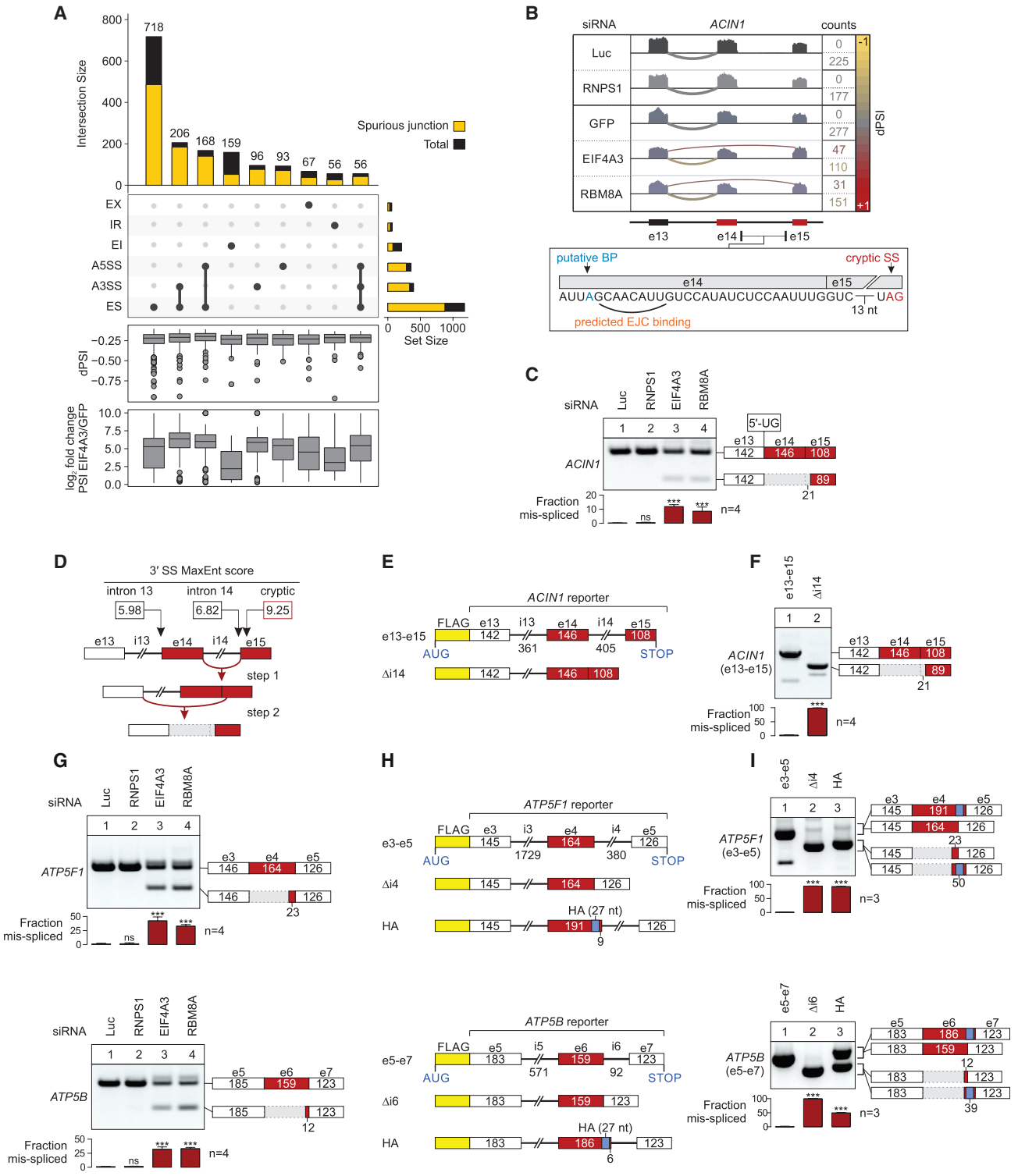


Figure 5. Inability to Deposit EJCs on mRNA Results in the Usage of Aberrant 3′ Splice Sites

(A) Classification of selected alternatively spliced junctions upon EIF4A3 knockdown. The quantity of spurious junctions for each class is shown. The change in junction usage (delta percent spliced in, dPSI) and PSI fold change for each class are shown as boxplots on the bottom. The full classification plot is shown in Figure S6A.

(B) Sashimi-plots of multiple RNA-Seq datasets (Luc and RNPS1 versus GFP, EIF4A3, and RBM8A) of *ACIN1* exon 14 skipping. The predicted EJC binding site in relation to the cryptic splice site (SS) and putative branchpoints (BP) is indicated.

(legend continued on next page)

EJC-dependent splicing regulation, we analyzed the *ACIN1* transcript, in which a cryptic 3′SS was used for splicing upon EIF4A3 or RBM8A, but not RNPS1 depletion (Figures 5B and 5C). We reasoned that a splicing order in which splicing of intron 14 precedes splicing of intron 13 could explain the observed splicing pattern (Figure 5D). Splicing of intron 14 generates a composite exonic 3′SS spanning the junction between exons 14 and 15, which led to exon skipping if used together with the 5′SS of intron 13. Consistent with our theory, an *ACIN1* reporter lacking intron 14 (Figure 5E) produced almost exclusively the skipped transcript, whereas a reporter construct with two introns was normally spliced (Figure 5F). The same order of splicing events was also observed for *ATP5B* and *ATP5F1* transcripts, for which mis-splicing is enabled due to the activation of a cryptic 3′ splice acceptor site (Figures 5G–5I). Similarly, splicing of the downstream intron precedes splicing of the upstream intron in the *CIAO1* mRNA (Figure S6B–S6J). As a result, the EJC helps to suppress two cryptic 3′SS downstream of the exon-exon junction by RNPS1-dependent and -independent mechanisms. For all transcripts analyzed in detail, the predicted EJC binding sites coincide with putative branchpoints or polypyrimidine tracts of the EJC-dependent alternative 3′SS. Furthermore, the EJC-suppressed alternative 3′SS are mainly clustered at EJC binding sites or at the 5′ end of the following exon (Figure S5Q). This finding suggests that the RNPS1-independent EJC function is due to direct masking of important splice-regulatory sequences. This hypothesis was supported by the insertion of hemagglutinin (HA) tag sequences in the *ATP5F1* and *ATP5B* reporter, which moved the alternative 3′SS sufficiently upstream of EJC deposition sites, thereby de-repressing and constitutively activating these SS (Figures 5H and 5I). Taken together, in addition to the PSAP-mediated splicing regulation, the deposition of EJCs prevents the loss of exonic sequences via usage of reconstituted or cryptic 3′SS.

EJCs Maintain Transcriptome Integrity and Cellular Survival

The importance of EJC and RNPS1-dependent SS protection is underscored by the observation that RBM8A-depleted cells and to a lesser extend RNPS1-depleted cells showed reduced proliferation (Figure 6A), presumably caused by the activation of several signaling cascades related to stress and apoptosis (Figure 6B). As the EJC is involved in a multitude of gene expression processes—for example, mRNA quality control by NMD—the cellular stress cannot be solely attributed to the uncontrolled splicing upon EJC knockdown. However, compared to the combined depletion of RNPS1 and SMG1, which is a central component of the NMD machinery, downregulation of RBM8A caused a more severe stress phenotype (Figure 6B). We suspected that normal EJC deposition and concomitant cryptic SS repression maintains proper transcriptome integrity and is therefore required for cellular fitness. Indeed, increased production of several shorter mRNA variants was observed in EJC knockdowns, when studying genes that harbor more than one EJC-dependent cryptic SS (Figures 6C–6F). We reasoned that this shift from productive to non-productive splicing and therefore the loss of intact, full-length transcripts might have a significant effect on cellular survival. In addition, we searched for mis-splicing of genes essential for cellular fitness by comparing the change of junction usage of EIF4A3-dependent spurious splice events with the fitness score of the respective gene, determined by a high-resolution CRISPR-screen (Figure 6G; Hart et al., 2015). In total, we found 184 mis-spliced genes that were classified as essential for survival and proliferation. Although many of the splice changes affect only one exon, we found many essential genes that produce substantial amounts of mis-spliced mRNAs, which are expected to encode for non-functional proteins. One example is the transcript of the splicing and DNA damage repair component *PRPF19*, for which exon 15 skipping upon EJC depletion leads to the deletion of two WD40 domains (Figure 6H), rendering the protein incapable of supporting proper DNA damage response (Maréchal et al., 2014). Another essential target is uridine monophosphate synthetase (*UMPS*), which is required for the UMP biosynthesis pathway. *UMPS* transcripts lacking exon 2 due to exon skipping generate an early frameshift and thus lead to the expression of a truncated protein isoform (Figure 6I). Therefore, the deposition of EJC on nascent mRNAs has an instant protective effect on gene expression, maintains the expression of many essential genes, and presumably represents an essential function of the EJC.

DISCUSSION

The human genome contains many cryptic SS that are usually used only at a very low frequency despite having similar sequences as canonical SS (Buratti et al., 2011). It is assumed that cryptic SS are suppressed by stronger SS in their vicinity and are only activated by mutations of nearby authentic SS (Roca et al., 2003). However, our data suggest that SS selection is not solely the result of SS competing for the splicing machinery. We find that many cryptic and reconstituted SS are efficiently silenced under normal conditions by the deposition of an EJC on adjacent exon-exon junctions. Furthermore, we uncover several different mechanisms by which EJCs globally inhibit the use of nearby irregular 5′ and 3′SS (Figure 7A). Interestingly, a parallel study reports a similar observation that the splicing of hundreds of putative RS exons is regulated by the EJC and associated factors (Blazquez et al., 2018). Because the EJC itself is

(C) RT-PCR analysis of *ACIN1* exon 14 skipping with RNA from HeLa cells transfected with the indicated siRNA.
(D) Scheme indicating the *ACIN1* re-splicing mechanism.
(E and F) *ACIN1* minigene reporter constructs (E) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (F).
(G) *ATP5F1* and *ATP5B* RT-PCR of cDNA samples obtained from HeLa cells transfected with the indicated siRNA.
(H and I) *ATP5F1* and *ATP5B* minigene reporter constructs (H) were expressed in HeLa Tet-Off cells and analyzed via RT-PCR (I). Insertion of 27 nt HA sequences in exon 4 (*ATP5F1*) or exon 6 (*ATP5B*) is schematically depicted.
All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.
See also Figures S5 and S6 and Table S2.

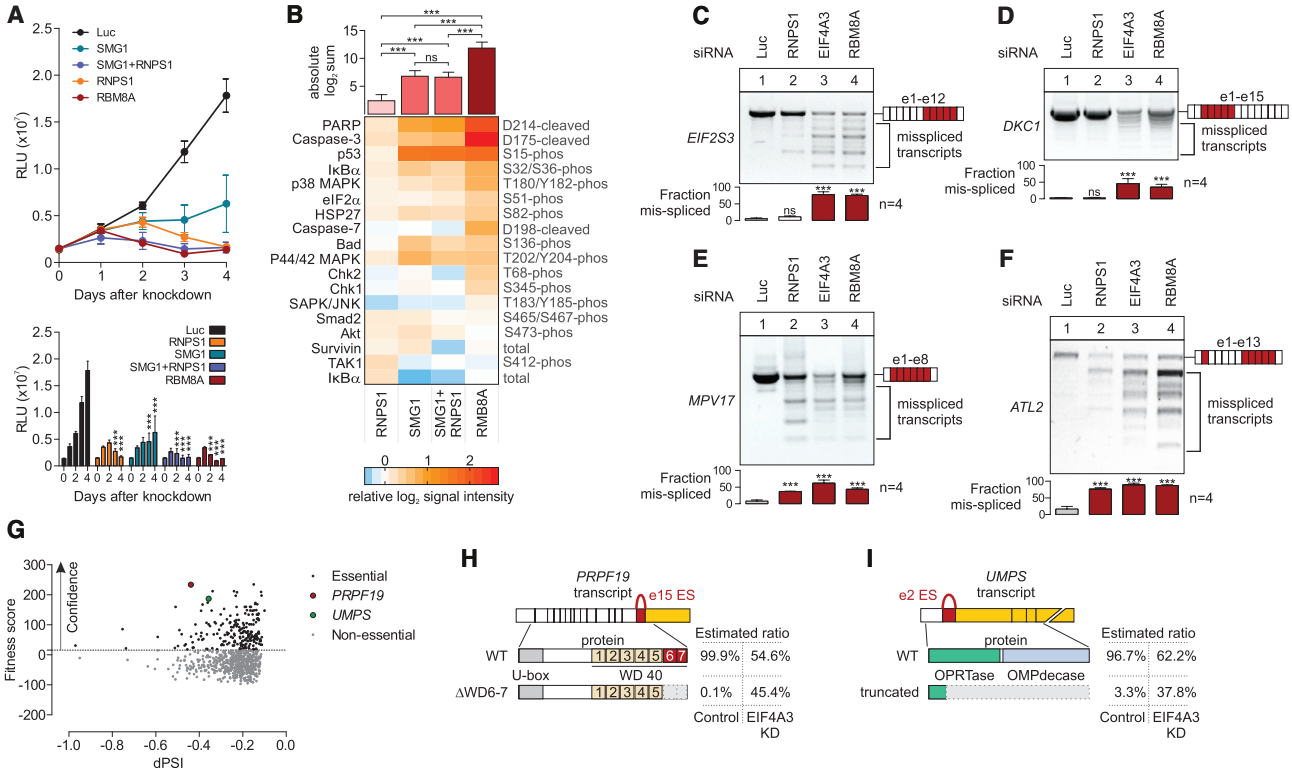


Figure 6. RNPS1 and EJC Depletion Leads to Transcript Disintegration and Cellular Stress

(A) Proliferation of HeLa cells was measured after knockdown using the indicated siRNA. n = 3.

(B) Induction of stress pathways upon transfection of HeLa cells with the indicated siRNA. The indicated signaling molecules were detected via target-specific capture antibodies in a sandwich immunoassay. n = 3.

(C–F) Examples for accumulated mis-spliced *EIF2S3* (C), *DKC1* (D), *MPV17* (E), and *ATL2* (F) transcripts upon EJC depletion, analyzed by RT-PCR.

(G) Plot of spurious junctions identified in EIF4A3 knockdown RNA-Seq data, comparing the log Bayes factor (BF) gene fitness score against the change in junction usage (dPSI). More positive scores increase the confidence in the essentiality of the gene. Individual targets are highlighted.

(H and I) Scheme of highlighted targets in (G), depicting the transcript architecture, the strongest alternative splicing event, the expressed protein variants, and the ratio of mis-splicing in control or EIF4A3 knockdown RNA-Seq data.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

deposited on the RNA during splicing, this mechanism functions as a positive feedback loop to reinforce authentic SS and establishes a hierarchy of preferential SS usage.

Mechanism and Consequences of 5′SS Suppression

The position of the EJC at the 3′ end of an exon is well suited to oversee the 5′ end of the next exon and to suppress exonic SS. Our work establishes that the EJC recruits the PSAP complex to exert an exon inclusion effect that decreases proportionally with the distance. Consequently, PSAP functions mainly in the vicinity of previously spliced introns, so that cryptic SS usage within already-ligated exons is prevented. In contrast to other splice factors such as SR proteins, for which several binding sequences can be present in an exon (Long and Caceres, 2009), EJCs recruit PSAP complexes in a splicing-dependent but sequence-independent manner to a single site at the 3′ end of an exon. Despite its limited binding potential/possibilities, the efficiency of PSAPs cryptic splice site suppression is remarkable. In many cases, EJC-regulated cryptic splices are barely used under normal conditions. While the knockdown of EJC components often causes only partial mis-splicing of endogenous transcripts (presumably due to incomplete knockdown of the proteins), we observe almost 100% splicing defects in our reporter constructs when EJC binding was prevented by intron deletions. Interestingly, there are also similarities between PSAP complex and SR proteins. For example, we observe a distance-dependent exonization effect by the PSAP complex, as has been described for SR proteins (Graveley et al., 1998). While several components of an EJC-dependent exon definition complex are now characterized, more work will be needed to determine the exact mechanism of 5′SS suppression. It will be of particular importance to uncover the process by which the PSAP complex guides U1 snRNP binding to 5′SS.

Mechanism and Consequences of 3′SS Protection

In addition to the PSAP-dependent 5′SS suppression, we observed a different, merely steric mechanism in the control of 3′SS. We conclude that the binding of the EJC core prevents the correct recognition of the 3′SS by the spliceosome, likely because factors such as U2AF1/2 are not able to bind to the

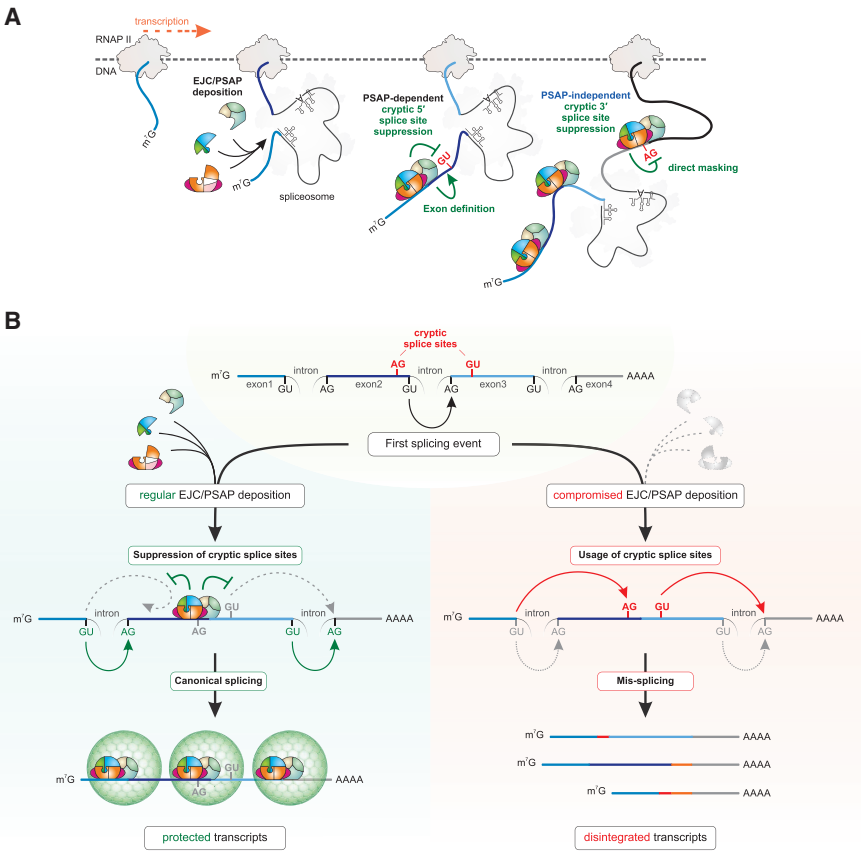


Figure 7. EJC and RNPS1 Protect Spliced Transcripts from the Usage of Cryptic Splice Sites

(A) Mechanism of splice site suppression by the EJC and the RNPS1-containing PSAP complex.

(B) Deposition of EJCs prevents loss of exonic sequences by masking and suppressing cryptic splice sites in the vicinity, consequently enforcing correct splicing hierarchy. The inability to assemble EJCs on spliced transcripts results in the activation of cryptic SS, leading to mis-splicing events and loss of exonic sequences.

RNA if an EJC occupies their genuine binding site. A similar interference with the assembly of the early spliceosome has been reported for a diverse group of repressors of 3′SS usage, including PTB or Sxl (Izquierdo et al., 2005; Valcárcel et al., 1993). Many of these factors compete with U2AF for binding to the pyrimidine tract, and therefore the inhibitory potential of the repressor depends on its binding strength to the 3′SS (Sohail and Xie, 2015). In contrast, the EJC has a clear competitive advantage over U2AF because its co-spliceosomal deposition ensures that cryptic or reconstituted SS are nearly immediately rendered inaccessible. Furthermore, the strong EJC binding to mRNA is usually only resolved by a translating ribosome in the cytoplasm (Gehring et al., 2009b). This advantage explains the surprisingly robust effect of the EJC on the suppression of 3′SS.

Significance of the EJC for Gene Expression

It is evident that the complex and heterogeneous architecture of the human transcriptome demands a mechanism that marks the position of already spliced introns and prevents re-splicing in their vicinity. The EJC meets this demand by masking and suppressing aberrant SS and thereby protecting the bound transcript from unintentional loss of exonic sequences (Figure 7B). Conceptually, this protective effect of the EJC is comparable to the telescripting mechanism in which the U1 snRNP prevents premature transcription termination by cleavage and polyadenylation of mRNAs in order to ensure transcriptome integrity (Berg et al., 2012). Similar to U1 depletion resulting in mRNA shortening, EJC depletion results in transcript disintegration by de-repression of single or multiple SS in a given transcript (Figure 7B). Mechanistically, the EJC-dependent re-splicing events share characteristics with reported multi-step splice processes such as recursive splicing or intra-splicing (Parra et al., 2008; Sibley et al., 2016). Functionally, however, the EJC-related re-splicing events reported here result primarily in destructive splice patterns, whereas canonical recursive splicing is defined as a productive mechanism resulting in the correct excision of introns.

We speculate that reduced expression of the EJC core components EIF4A3 and RBM8A can lead to significant changes in the transcriptome, explaining why the EJC is important for embryonic development (Mao et al., 2016) and how its misregulation can cause serious human diseases (Albers et al., 2012; Favaro et al., 2014). In sensitive cells or tissues, these changes in gene expression will affect cellular fitness and eventually have a negative impact on tissue differentiation and maintenance. Hence, our work uncovers EJCs as essential components of a splice-regulatory pathway that safeguard transcriptome integrity and protect mRNAs against disruptive splicing events.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
- METHOD DETAILS
  - Stable Cell Lines and Plasmids
  - *In Vitro* Transcription, *In Vitro* Splicing and RNP Immunoprecipitation
  - Co-immunoprecipitation
  - siRNA Transfections
  - Transient Plasmid Transfections

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and three tables and can be found with this article online at https://doi.org/10.1016/j.molcel.2018.08.030.

ACKNOWLEDGMENTS

We thank members of the Gehring Lab for discussions and reading of the manuscript. We also thank Marek Franitza and Christian Becker (Cologne Center for Genomics, CCG) for preparing the sequencing libraries and operating the sequencer. We are grateful to Thomas Wiehe and Peter Heger (Institute for Genetics, Cologne) for sharing computational infrastructure. We also thank Jernej Ule for sharing their finding on the EJC-dependent regulation of recursive splicing at the onset of our study. This work was supported by grants from the DST to K.K.S. (ECR/2015/000166) and the Deutsche Forschungsgemeinschaft to A.-L.S. (STE 2509/2-1), C.D. (DI 1501/8-1), and N.H.G (GE 2014/6-1). V.B. was funded under the Institutional Strategy of the University of Cologne within the German Excellence Initiative. N.H.G. acknowledges support by a Heisenberg fellowship (GE 2014/5-1 and GE 2014/7-1) from the Deutsche Forschungsgemeinschaft. C.D. and T.B.B. were kindly supported by the Klaus Tschira Stiftung gGmbH (00.219.2013).

AUTHOR CONTRIBUTIONS

Conceptualization, N.H.G., V.B., and L.B.; Methodology, N.H.G., V.B., A.-L.S., K.K.S., and J.V.G.; Software, T.B.B., C.D., and V.B.; Investigation, V.B., A.-L.S., K.K.S., E.G., and J.A.; Resources and Data Curation, T.B.B. and C.D.; Writing – Original Draft, Review & Editing, N.H.G., V.B., A.-L.S., T.B.B., C.D., J.V.G., and K.K.S.; Visualization, V.B.; Supervision and Funding Acquisition, N.H.G. and C.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 26, 2018
Revised: July 24, 2018
Accepted: August 20, 2018
Published: November 1, 2018

REFERENCES

Albers, C.A., Paul, D.S., Schulze, H., Freson, K., Stephens, J.C., Smethurst, P.A., Jolley, J.D., Cvejic, A., Kostadima, M., Bertone, P., et al. (2012). Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. Nat Genet *44*, 435–439, S431–432.

Alexandrov, A., Colognori, D., Shu, M.D., and Steitz, J.A. (2012). Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. Proc. Natl. Acad. Sci. USA *109*, 21313–21318.

Andersen, C.B., Ballut, L., Johansen, J.S., Chamieh, H., Nielsen, K.H., Oliveira, C.L., Pedersen, J.S., Séraphin, B., Le Hir, H., and Andersen, G.R. (2006). Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science *313*, 1968–1972.

Ashton-Beaucage, D., Udell, C.M., Lavoie, H., Baril, C., Lefrançois, M., Chagnon, P., Gendron, P., Caron-Lizotte, O., Bonneil, E., Thibault, P., and Therrien, M. (2010). The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. Cell *143*, 251–262.

Barbosa, I., Haque, N., Fiorini, F., Barrandon, C., Tomasetto, C., Blanchette, M., and Le Hir, H. (2012). Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. Nat. Struct. Mol. Biol. *19*, 983–990.

Berg, M.G., Singh, L.N., Younis, I., Liu, Q., Pinto, A.M., Kaida, D., Zhang, Z., Cho, S., Sherrill-Mix, S., Wan, L., and Dreyfuss, G. (2012). U1 snRNP determines mRNA length and regulates isoform expression. Cell *150*, 53–64.

Blazquez, L., Emmett, W., Faraway, R., Pineda, J.M.B., Bajew, S., Gohr, A., Haberman, N., Sibley, C.R., Bradley, R.K., Irimia, M., and Ule, J. (2018). Exon junction complex shapes the transcriptome by repressing recursive splicing. Mol. Cell *72*, this issue, 496–509.

Boehm, V., and Gehring, N.H. (2016). Exon junction complexes: supervising the gene expression assembly line. Trends Genet. *32*, 724–735.

Bono, F., Ebert, J., Lorentzen, E., and Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell *126*, 713–725.

Buratti, E., Chivers, M., Hwang, G., and Vorechovsky, I. (2011). DBASS3 and DBASS5: databases of aberrant 3′- and 5′-splice sites. Nucleic Acids Res. *39*, D86–D91.

Cáceres, E.F., and Hurst, L.D. (2013). The evolution, impact and properties of exonic splice enhancers. Genome Biol. *14*, R143.

Corvelo, A., Hallegger, M., Smith, C.W., and Eyras, E. (2010). Genome-wide association between branch point properties and alternative splicing. PLoS Comput. Biol. *6*, e1001016.

Dale, R.K., Pedersen, B.S., and Quinlan, A.R. (2011). Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics *27*, 3423–3424.

Desmet, F.O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. *37*, e67.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dodt, M., Roehr, J.T., Ahmed, R., and Dieterich, C. (2012). FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. Biology (Basel) *1*, 895–905.

Favaro, F.P., Alvizi, L., Zechi-Ceide, R.M., Bertola, D., Felix, T.M., de Souza, J., Raskin, S., Twigg, S.R., Weiner, A.M., Armas, P., et al. (2014). A noncoding expansion in EIF4A3 causes Richieri-Costa-Pereira syndrome, a craniofacial disorder associated with limb defects. Am. J. Hum. Genet. *94*, 120–128.

Fukumura, K., Inoue, K., and Mayeda, A. (2018). Splicing activator RNPS1 suppresses errors in pre-mRNA splicing: A key factor for mRNA quality control. Biochem. Biophys. Res. Commun. *496*, 921–926.

Gehring, N.H., Lamprinaki, S., Hentze, M.W., and Kulozik, A.E. (2009a). The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. PLoS Biol. *7*, e1000120.

Gehring, N.H., Lamprinaki, S., Kulozik, A.E., and Hentze, M.W. (2009b). Disassembly of exon junction complexes by PYM. Cell *137*, 536–548.

Graveley, B.R., Hertel, K.J., and Maniatis, T. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. EMBO J. *17*, 6747–6756.

Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. Biochem. J. *430*, 379–392.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K.R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. Cell *163*, 1515–1526.

Hauer, C., Sieber, J., Schwarzl, T., Hollerer, I., Curk, T., Alleaume, A.M., Hentze, M.W., and Kulozik, A.E. (2016). Exon junction complexes show a distributional bias toward alternatively spliced mRNAs and against mRNAs coding for ribosomal proteins. Cell Rep. *16*, 1588–1603.

Hayashi, R., Handler, D., Ish-Horowicz, D., and Brennecke, J. (2014). The exon junction complex is required for definition and excision of neighboring introns in Drosophila. Genes Dev. *28*, 1772–1785.

Ibrahim, E.C., Schaal, T.D., Hertel, K.J., Reed, R., and Maniatis, T. (2005). Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proc. Natl. Acad. Sci. USA *102*, 5002–5007.

Izquierdo, J.M., Majós, N., Bonnal, S., Martínez, C., Castelo, R., Guigó, R., Bilbao, D., and Valcárcel, J. (2005). Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. Mol. Cell *19*, 475–484.

Katz, Y., Wang, E.T., Silterra, J., Schwartz, S., Wong, B., Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., Airoldi, E.M., and Burge, C.B. (2015). Quantitative visualization of alternative exon expression from RNA-seq data. Bioinformatics *31*, 2400–2402.

Kornblihtt, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nat. Rev. Mol. Cell Biol. *14*, 153–165.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Le Hir, H., Saulière, J., and Wang, Z. (2016). The exon junction complex as a node of post-transcriptional networks. Nat. Rev. Mol. Cell Biol. *17*, 41–54.

Lee, Y., and Rio, D.C. (2015). Mechanisms and regulation of alternative Pre-mRNA splicing. Annu. Rev. Biochem. *84*, 291–323.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. IEEE Trans. Vis. Comput. Graph. *20*, 1983–1992.

Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. Biochem. J. *417*, 15–27.

Malone, C.D., Mestdagh, C., Akhtar, J., Kreim, N., Deinhard, P., Sachidanandam, R., Treisman, J., and Roignant, J.Y. (2014). The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript. Genes Dev. *28*, 1786–1799.

Mao, H., McMahon, J.J., Tsai, Y.H., Wang, Z., and Silver, D.L. (2016). Haploinsufficiency for core exon junction complex components disrupts embryonic neurogenesis and causes p53-mediated microcephaly. PLoS Genet. *12*, e1006282.

Maréchal, A., Li, J.M., Ji, X.Y., Wu, C.S., Yazinski, S.A., Nguyen, H.D., Liu, S., Jiménez, A.E., Jin, J., and Zou, L. (2014). PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry. Mol. Cell *53*, 235–246.

Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M.Q., Gardiner, E.M., and Krainer, A.R. (1999). Purification and characterization of human RNPS1: a general activator of pre-mRNA splicing. EMBO J. *18*, 4560–4570.

Murachelli, A.G., Ebert, J., Basquin, C., Le Hir, H., and Conti, E. (2012). The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. Nat. Struct. Mol. Biol. *19*, 378–386.

Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. Nature *463*, 457–463.

Papasaikas, P., and Valcárcel, J. (2016). The spliceosome: the ultimate RNA chaperone and sculptor. Trends Biochem. Sci. *41*, 33–45.

Parra, M.K., Tan, J.S., Mohandas, N., and Conboy, J.G. (2008). Intrasplicing coordinates alternative first exons with alternative splicing in the protein 4.1R gene. EMBO J. *27*, 122–131.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. *44* (W1), W83–W89.

Roca, X., Sachidanandam, R., and Krainer, A.R. (2003). Intrinsic differences between authentic and cryptic 5′ splice sites. Nucleic Acids Res. *31*, 6321–6333.

Rodgers, P., Stapleton, G., Flower, J., and Howse, J. (2014). Drawing area-proportional Euler diagrams representing up to three sets. IEEE Trans. Vis. Comput. Graph. *20*, 56–69.

Rodor, J., Pan, Q., Blencowe, B.J., Eyras, E., and Cáceres, J.F. (2016). The RNA-binding profile of Acinus, a peripheral component of the exon junction complex, reveals its role in splicing regulation. RNA *22*, 1411–1426.

Roignant, J.Y., and Treisman, J.E. (2010). Exon junction complex subunits are required to splice Drosophila MAP kinase, a large heterochromatic gene. Cell *143*, 238–250.

Sakashita, E., Tatsumi, S., Werner, D., Endo, H., and Mayeda, A. (2004). Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo. Mol. Cell. Biol. *24*, 1174–1187.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. *3*, 1101–1108.

Sibley, C.R., Blazquez, L., and Ule, J. (2016). Lessons from non-canonical splicing. Nat. Rev. Genet. *17*, 407–421.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell *151*, 750–764.

Sohail, M., and Xie, J. (2015). Diverse regulation of 3′ splice site usage. Cell. Mol. Life Sci. *72*, 4771–4793.

Steckelberg, A.L., and Gehring, N.H. (2014). Studying the composition of mRNPs in vitro using splicing-competent cell extracts. Methods *65*, 342–349.

Steckelberg, A.L., Boehm, V., Gromadzka, A.M., and Gehring, N.H. (2012). CWC22 connects pre-mRNA splicing and exon junction complex assembly. Cell Rep. *2*, 454–461.

Tapial, J., Ha, K.C.H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y., et al. (2017). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. Genome Res. *27*, 1759–1768.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. *7*, 562–578.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Valcárcel, J., Singh, R., Zamore, P.D., and Green, M.R. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. Nature *362*, 171–175.

Vaquero-Garcia, J., Barrera, A., Gazzara, M.R., González-Vallinas, J., Lahens, N.F., Hogenesch, J.B., Lynch, K.W., and Barash, Y. (2016). A new view of transcriptome complexity and regulation through the lens of local splicing variations. eLife *5*, e11752.

Wang, Z., Murigneux, V., and Le Hir, H. (2014). Transcriptome-wide modulation of splicing by the exon junction complex. Genome Biol. *15*, 551.

Wang, Z., Ballut, L., Barbosa, I., and Le Hir, H. (2018). Exon Junction Complexes can have distinct functional flavours to regulate specific splicing events. Sci. Rep. *8*, 9509.

Woodward, L.A., Mabin, J.W., Gangras, P., and Singh, G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. Wiley Interdiscip. Rev. RNA *8*. Published online December 23, 2016. https://doi.org/10.1002/wrna.1411.

Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J. Comput. Biol. *11*, 377–394.

Young, L., Sung, J., Stacey, G., and Masters, J.R. (2010). Detection of Mycoplasma in cell cultures. Nat. Protoc. *5*, 929–934.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-FLAG	Sigma-Aldrich	Cat# F7425; RRID:AB_439687
Mouse monoclonal anti-beta-Actin	Sigma-Aldrich	Cat# A5441; RRID:AB_476744
Mouse monoclonal anti-alpha-Tubulin	Sigma-Aldrich	Cat# T6074; RRID:AB_477582
Rabbit polyclonal anti-V5	QED Bioscience	Cat# 18870-01; RRID:AB_215340
Rabbit polyclonal anti-GFP	Abcam	Cat# ab290; RRID:AB_303395
Rabbit polyclonal anti-SAP18	Abcam	Cat# ab31748; RRID:AB_777829
Rabbit polyclonal anti-EIF4A3; raised against an N-terminal peptide of EIF4A3	GenScript	N/A
Peroxidase-AffiniPure Goat Anti-Rabbit	Jackson ImmunoResearch Laboratories	Cat# 111-035-006; RRID:AB_2337936
Peroxidase-AffiniPure Goat Anti-Mouse	Jackson ImmunoResearch Laboratories	Cat# 115-035-003; RRID:AB_10015289
Chemicals, Peptides, and Recombinant Proteins		
Doxycycline	Takara Bio/Clontech	Cat# 8634-1
Cumate 10,000x solution	System Biosciences	Cat# QM150A-1
Ribo m7G Cap Analog	Promega	Cat# P1711
3′-O-Me-m7G(5′)ppp(5′)G RNA Cap Structure Analog	New England Biolabs	Cat# S1411S
α-32P-GTP	Hartmann Analytic	Cat# FP808
HeLa nuclear extract	CIL Biotech	Cat# CC-01-20-50
Puromycin Dihydrochlorid	Carl Roth	Cat# 0240.1
Hygromycin B Gold	InvivoGen	Cat# ant-hg-1
RNasin	Promega	Cat# N2111
Critical Commercial Assays		
Lipofectamine RNAiMAX Transfection Reagent	Thermo Fisher Scientific	Cat# 13778150
EZview Red ANTI-FLAG M2 Affinity Gel	Sigma-Aldrich	Cat# F2426-1ML
Anti-FLAG M2 Magnetic Beads	Sigma-Aldrich	Cat# M8823-5ML
Amersham ECL Prime Western Blotting Detection Reagent	GE Healthcare	Cat# RPN2236
jetPRIME Transfection Reagent	Polyplus Transfection	Cat# 114-07
peqGOLD TriFast	VWR	Cat# 30-2010
ProtoScript II Reverse Transcriptase	New England Biolabs	Cat# M0368S
GoScript Reverse Transcriptase	Promega	Cat# A5003
GoScript Reverse Transcription Mix, Oligo(dT)	Promega	Cat# A2791
GoTaq Green Master Mix	Promega	Cat# M7123
GoTaq qPCR Master Mix	Promega	Cat# A6001
PathScan Stress and Apoptosis Signaling Antibody Array Kit	Cell Signaling Technology	Cat# 12856
CellTiter-Glo Luminescent Cell Viability Assay	Promega	Cat# G7570
Deposited Data		
Raw and analyzed data: RNA-seq	This paper	ArrayExpress: E-MTAB-6564
Raw and analyzed data: RNA-seq	Wang et al., 2014	GEO: GSE63091
ENSEMBL Human genome assembly	N/A	ftp://ftp.ensembl.org/pub/release-90/fasta/homo_sapiens/dna/RRID:SCR_002344
Mendeley dataset of original Data	This paper	https://doi.org/10.17632/wt7ybwz82g.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
Human: Flp-In-T-REx-293	Thermo Fisher Scientific	Cat# R78007; RRID:CVCL_U427
Human: HeLa Tet-Off	Clontech	Cat# 631156; RRID:CVCL_V352
Human: HeLa Flp-In T-REx	Elena Dobrikova and Matthias Gromeier, Duke University Medical Center	N/A
Oligonucleotides		
siRNA targeting sequence: Luciferase: CGUACGCGGAAUACUUCGA	Integrated DNA Technologies	N/A
siRNA targeting sequence: RNPS1_3: AAGGAAGACCAGUAGGAAA	Integrated DNA Technologies	N/A
siRNA targeting sequence: RNPS1_A: GGAUCGCUCAAAAGAUAAA	Integrated DNA Technologies	N/A
siRNA targeting sequence: RBM8A: UUCGCAGAAUAUGGGGAAA	Integrated DNA Technologies	N/A
siRNA targeting sequence: EIF4A3: AGACAUGACUAAAGUGGAA	Integrated DNA Technologies	N/A
siRNA targeting sequence: ACIN1_A: CUGCAGAGCAUGAAGUAAAUU	Integrated DNA Technologies	N/A
siRNA targeting sequence: ACIN1_13.1: UCAGUAUCACCACUGAAUCACUAAA	Integrated DNA Technologies	N/A
siRNA targeting sequence: ACIN1_13.2: CAAGUGAAAACAGACCUGAAAAUGA	Integrated DNA Technologies	N/A
siRNA targeting sequence: PNN_1: GAUUUCUUGAUAAAAAAGGAUUACC	Integrated DNA Technologies	N/A
siRNA targeting sequence: PNN_2: AUACUUCAGGACUAGAAAGAAGUCA	Integrated DNA Technologies	N/A
siRNA targeting sequence: SAP18_1: UCUGCAACCUUGGAUAAAUUAUUTC	Integrated DNA Technologies	N/A
siRNA targeting sequence: SAP18_2: GGAUGCAACCUUGAAAGAACUGACA	Integrated DNA Technologies	N/A
siRNA targeting sequence: SMG1: GUGUAUGUGCGCCAAAGUATT	Integrated DNA Technologies	N/A
siRNA targeting sequence: SRSF1: CCAAGGACAUUGAGGACGU	Integrated DNA Technologies	N/A
siRNA targeting sequence: SRSF2: AAUCCAGGUCGCGAUCGAA	Integrated DNA Technologies	N/A
siRNA targeting sequence: SRSF11_13.1: GUUGAUGUACAGUUUCUGAAAUUGT	Integrated DNA Technologies	N/A
siRNA targeting sequence: SRSF11_13.2: UACUGUUGAUCCCAAGUUGAAUCAT	Integrated DNA Technologies	N/A
Primers for RT-PCR and reporters see Table S3	This paper	N/A
Recombinant DNA		
pcDNA5/FRT/TO Mammalian Expression Vector	Thermo Fisher Scientific	Cat# V652020
pCI-neo Mammalian Expression Vector	Promega	Cat# E1841
PB-Cuo-MCS-IRES-GFP-EF1α-CymR-Puro Inducible cDNA Cloning and Expression Vector	System Biosciences	Cat# PBQM812A-1
Software and Algorithms		
Flexbar 3.0	Dodt et al., 2012	https://github.com/seqan/flexbar
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
STAR 2.5.3a	Dobin et al., 2013	https://github.com/alexdobin/STAR

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cufflinks 2.2.1	Trapnell et al., 2013	http://cole-trapnell-lab.github.io/cufflinks
Cummerbund	Trapnell et al., 2012	http://compbio.mit.edu/cummeRbund/
Majiq 1.0.6	Vaquero-Garcia et al., 2016	https://majiq.biociphers.org/
StringTie 1.3.3b	Pertea et al., 2015	https://github.com/gpertea/stringtie
Gffcompare 0.10.1	Geo Pertea	https://github.com/gpertea/gffcompare
GraphPad Prism 5.03	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
Superheat R package	arXiv:1512.01524 [stat.AP]	https://github.com/rlbarter/superheat
UpSet R package	Lex et al., 2014	http://caleydo.org/tools/upset/
Euler3 Applet	Rodgers et al., 2014	http://www.eulerdiagrams.com/Euler3.html
Human Splicing Finder online tool	Desmet et al., 2009	http://www.umd.be/HSF3/
SVM-BPfinder	Corvelo et al., 2010	http://regulatorygenomics.upf.edu/Software/SVM_BP/
Integrative Genomics Viewer (Version 2.3.98)	Thorvaldsdóttir et al., 2013	http://software.broadinstitute.org/software/igv/RRID:SCR_011793
Image Lab software	Bio-Rad	SOFT-LIT-170-9690-ILSPC
ImageQuant TL software	GE Healthcare	Cat# 29000605

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Niels H. Gehring (ngehring@uni-koeln.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

HEK293 Flp-In T-REx (human; sex: female; Thermo Fisher Scientific; RRID:CVCL_U427), HeLa Flp-In T-REx (human; sex: female; established by Elena Dobrikova and Matthias Gromeier, Duke University Medical Center) and HeLa Tet-Off (human; sex: female; Clontech; RRID:CVCL_V352) cells were cultured in DMEM (GIBCO) supplemented with 9% fetal bovine serum (GIBCO) and 1x Penicillin Streptomycin (GIBCO). All cells were cultivated at 37°C and 5% CO2 in a humidified incubator.

METHOD DETAILS

Stable Cell Lines and Plasmids

Mammalian expression constructs for *in vitro* splicing assays were inserted into the pCI-neo vector (Promega) with an N-terminal FLAG tag. MINX and MINX Δintron *in vitro* splice substrates were described previously (Gehring et al., 2009a; Gehring et al., 2009b). The point and deletion mutants of RNPS1 were PCR amplified and inserted into pCI-neo-FLAG. Accordingly, GST, CWC22 WT and CWC22 NK171DE mutant (described in (Steckelberg et al., 2012)) were cloned into pCI-neo-FLAG. For transient tethering assays, the constructs were subcloned into pCI-neo containing an N-terminal MS2V5 tag. For generating stable tethering cell lines, the constructs together with the MS2V5 tag were inserted in the cumate-inducible PB-CuO-MCS-IRES-GFP-EF1-CymR-Puro vector (System Biosciences). All reporter constructs were PCR amplified from either HeLa cDNA or genomic DNA and, if applicable, mutagenized by PCR. The 2xMS2-ΔESE binding sites were optimized via PCR to remove potential ESE sequences and inserted with varying spacers into the truncated exon 4 of the *HSD17B10* mini-gene. All minigene constructs were cloned in-frame with an N-terminal FLAG-tag into the pcDNA5/FRT/TO vector (Thermo Fisher Scientific). For generating stable RNPS1, SAP18 or PNN rescue cell lines, the expression constructs were cloned into the pcDNA5/FRT/TO vector containing an N-terminal FLAG-tag. To ensure robust expression in rescue assays, selected RNPS1 or PNN constructs were also cloned into the pcDNA5/FRT/TO vector containing an N-terminal FLAG-emGFP-tag. Standard protocols were used to generate stable rescue or reporter HeLa Flp-In T-REx cell lines and positive clones were selected with 100-150 μg ml^{-1} hygromycin B (InvivoGen). Expression of stable cell lines was induced for minimum 24 h with 1 μg ml^{-1} doxycycline. Dual-inducible tethering cell lines were generated by integrating a PB-CuO-MS2V5 construct in stable Flp-In T-REx reporter cell lines. 2 μg ml^{-1} puromycin was used for the selection of positive clones. Expression of the dual-inducible cell lines was first induced with 30 μg ml^{-1} cumate

for 24h, followed by both 1 μg ml^{-1} doxycycline and 30 μg ml^{-1} cumate for another 24 h. Mycoplasma contamination was tested by PCR amplification of mycoplasma-specific genomic DNA (Young et al., 2010) or by using the Mycoplasmacheck service (Eurofins Genomics).

In Vitro Transcription, *In Vitro* Splicing and RNP Immunoprecipitation

In vitro transcription and *in vitro* splicing experiments were performed as described previously (Gehring et al., 2009a). The capped MINX transcripts were generated by run-off transcription with SP6 polymerase (Promega) in the presence of Ribo m7G Cap Analog (Promega) and α-^{32}P-GTP (Hartmann Analytic). *In vitro* splicing reactions were carried out in HeLa nuclear extracts (CIL Biotech) supplemented with HEK293 whole cell extracts expressing FLAG-tagged proteins. After splicing, immunoprecipitations were carried out with EZview Red ANTI-FLAG M2 Affinity Gel (Sigma-Aldrich) in EJC buffer (20 mM HEPES-KOH pH 7.9, 200 mM NaCl, 2 mM $MgCl_2$, 0.2% Triton X-100, 0.1% Nonidet-P40, 0.05% sodium deoxycholic acid). Subsequently, RNA was extracted from the bound proteins via peqGOLD TriFast (VWR) and resolved by denaturing PAGE. For detection of co-immunoprecipitated proteins, splicing reactions with 3′-O-Me-m7G(5′)ppp(5′)G RNA Cap Structure Analog (NEB)-capped but non-radioactively labeled MINX transcripts were performed. Immunoprecipitations were carried out with Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) in EJC buffer and co-immunoprecipitated proteins were eluted with SDS-sample buffer, separated by SDS-PAGE, and analyzed by immunoblotting.

Co-immunoprecipitation

FLAG-tagged proteins were expressed in stable HeLa Flp-In T-REx cells induced for 48 h and immunoprecipitated from 1 mg cell lysate (in 50 mM Tris [pH 7.2], 150 mM NaCl, 0.5% Triton X-100) for 2 h using Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich) in the presence or absence of RNase A (50 μg ml^{-1}). Beads were washed four times with lysis buffer and co-immunoprecipitated proteins were eluted with SDS-sample buffer, separated by SDS-PAGE, and analyzed by immunoblotting. Effectiveness of RNase A treatment was confirmed by ethidium bromide staining of total RNA isolated after immunoprecipitation.

siRNA Transfections

Cell lines were reverse transfected with 60 pmol siRNA per 2 × 10^5 cells using 2.5 μl Lipofectamine RNAiMAX (Thermo Fisher Scientific). Fresh medium was supplied 24 h after siRNA transfection. The used siRNA target sequences are listed in the Key Resources Table.

Transient Plasmid Transfections

2.8 × 10^5 HeLa Tet-Off cells were seeded in 6-well plates 24 h before transfection by calcium phosphate precipitation. For reporter assays, 0.5 μg of a mVenus expression plasmid, 1 μg reporter plasmid and 2 μg fill plasmid encoding for β-globin were transfected. For tethering experiments 0.5 μg of a mVenus expression plasmid, 0.5 μg reporter plasmid and 1 μg plasmid encoding for MS2V5-tagged proteins were transfected. Overexpression of proteins in HeLa Flp-In T-REx cells was performed using jetPRIME (Polyplus Transfection), co-transfecting 0.5 μg of a mVenus expression plasmid and 2 μg of plasmid encoding for FLAG-tagged proteins.

Immunoblot Analysis and Antibodies

SDS-polyacrylamide gel electrophoresis and immunoblot analysis were performed using protein samples derived from peqGOLD TriFast extractions, parallel transfection harvested with RIPA buffer or samples eluted from Anti-FLAG M2 magnetic beads. All antibodies (see Key Resources List) were used at 1:3000 dilutions in 50 mM Tris [pH 7.2], 150 mM NaCl with 0.2% Tween-20 and 5% skim milk powder. Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) in combination with the myECL Imager (Thermo Fisher Scientific) was used for visualization.

Reverse Transcription, End-Point and Quantitative RT-PCR

Cells were harvested with peqGOLD TriFast and RNA extracted according to the manufacturer's instructions. 1-4 μg of total RNA was reverse-transcribed in a 20 μl reaction volume with 10 μM VNN-$(dT)_{20}$ primer using the ProtoScript II Reverse Transcriptase (NEB), GoScript Reverse Transcriptase (Promega) or the GoScript Reverse Transcription Mix, Oligo(dT) (Promega). 0.5% of purified cDNA was used as template in end-point PCRs using the GoTaq Green Master Mix (Promega) and 0.2 μM final concentration of sense and antisense primer (see Table S3 for sequences). After 30 PCR cycles, the samples were resolved by electrophoresis on ethidium bromide-stained, 1% agarose TBE gels and visualized by trans-UV illumination using the Gel Doc XR+ (Bio-Rad). Representative gel images from at least three independent experiments are shown. Sanger sequencing of individual bands was performed using the service of Eurofins Genomics.

Quantitative RT-PCR were performed with the GoTaq qPCR Master Mix (Promega) using 0.5% of cDNA in 10 μl reactions, 0.2 μM final concentration of sense and antisense primer (see Table S3 for sequences), and the CFX96 Touch Real-Time PCR Detection System (Bio-Rad).

Pathscan and Cell Survival Assays

The PathScan Stress and Apoptosis Signaling Antibody Array Kit (Cell Signaling Technology) was used to detect cellular stress responses upon depletion of RNPS1, SMG1 or RBM8A. In brief, siRNA-mediated knockdowns of HeLa Flp-In T-REx were performed in 6-well plates as described above. After 3 days, the cells were harvested in 1x Cell Lysis Buffer and protein concentration was measured using the Bradford Protein Assay (Bio-Rad). Lysates were diluted to 0.5 mg ml^{-1} protein concentration and the sandwich immunoassay was performed according to the manufacturer's instructions. The cell lysate was incubated with nitrocellulose-coated glass slides on which target-specific capture antibodies have been spotted in duplicate. Biotinylated detection antibody cocktail in combination with streptavidin-conjugated HRP and LumiGLO Reagent are then used to visualize the bound detection antibody by chemiluminescence.

To measure cell proliferation and survival upon EJC component depletion, the CellTiter-Glo Luminescent Cell Viability Assay (Promega) was used. HeLa Flp-In T-REx were reverse transfected with siRNA as described above and 2000 cells were seeded in triplicates in 96-well plates suitable for sensitive luminescence measurements. 2 hours after seeding, the first time point was measured according to the manufacturer's instructions and subsequent measurements were performed in 24-hour intervals.

RNA-Seq

RNA-Seq analysis was carried out on normal or stable RNPS1-expressing HeLa Flp-In T-REx cells transfected with siRNAs targeting *RNPS1* or the negative control luciferase. Three biological replicates were analyzed for each sample. Total RNA was extracted with peqGOLD TriFast as described above. Ribosomal depletion and strand specific library preparation was carried out with the TruSeq R Stranded Total RNA LT (with Ribo-ZeroTM Human/Mouse/Rat) according to the manufacturer's instructions. After validation (Agilent 2200 TapeStation) and quantification (Invitrogen Qubit System) all 12 transcriptome libraries were pooled. The pool was quantified using the Peqlab KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection System and loaded on one lane of Illuminas HiSeq4000 sequencer with a 2 × 75bp protocol. The analysis produced 5.3 Gb/sample (4.8-6.6 Gb), corresponding to 35 Mread-pairs/sample. Basic read quality check was carried out using FastQC showing > 97.5% of Q30 bases (PF) and a mean quality score of 39,8.

Read Processing and Mapping

Adaptor sequences and low quality 3′ bases were removed with Flexbar 3.0. (Dodt et al., 2012). Short reads from the rRNA locus were subtracted by mapping against the 45S precursor (*Homo sapiens*, NR_046235.1) using Bowtie2 (Langmead and Salzberg, 2012). The remaining reads were aligned against the human genome (version 38, EnsEMBL 90 transcript annotations) using the STAR read aligner (2.5.3a)(Dobin et al., 2013)

Gene Expression Analysis

We employed the Cuffquant and Cuffdiff software (release 2.2.1) (Trapnell et al., 2013) to estimate gene expression abundance and differential gene expression for the EnsEMBL reference annotation. The R package Cummerbund (Trapnell et al., 2012) was subsequently used to inspect and visualize the results.

Local Splicing Variants Identification

We carried out local splicing variant detection using MAJIQ (1.0.6, without GC correction in the build step)(Vaquero-Garcia et al., 2016). To this end, we first produced a transcriptome annotation with StringTie (1.3.3b)(Pertea et al., 2015) for each RNA-Seq replicate on the control groups (Luc for RNPS1 versus Luc, GFP for EIF4A3 versus GFP) to compare with the knockdown conditions. Next, we combined the annotations using the merge command from StringTie, with the minimum isoform fraction (-f parameter) set to 0.5 to eliminate lowly transcribed isoform. Finally, we applied gffcompare (0.10.1) against the human genome annotation with the parameters –R (precision correction), -Q (sensitivity correction) and –M (discard single-exons transfrags). The stringent transcriptome annotation enabled us to contrast between conditions as it highlights differential exon usage.

The Voila tabular output was processed and analyzed with Python programming language (3.6). The dataset, which initially contains one local splice variations (LSV) per row, was expanded to provide one exon-exon junctions per row, enabling us to filter the exon-exon junction given the delta percent spliced in (dPSI) (< -0.1) and their posterior probability P(dPSI) (> 0.90). These cutoffs were applied to ensure the sensitive detection of alternative splicing events (dPSI < -0.1), while keeping only probable splicing events (P(dPSI) > 0.90). The dPSI between two conditions is calculated by estimating a posterior distribution for the change in the respective junction's relative inclusion level (Vaquero-Garcia et al., 2016). For all dPSI calculations the PSI of the RNPS1 or EIF4A3 knockdown sample was substracted from the PSI of the control (Luc or GFP). Furthermore, we used the bedtools intersect command (pybedtools 0.7.10, bedtools 2.26.0)(Dale et al., 2011; Quinlan and Hall, 2010) to extract the exon coordinates from the StringTie annotation overlapping the respective junction. This allowed us to re-classify all junctions based on calculating distances of each exon to junction connection. Furthermore, junctions were classified as spurious if the PSI in control samples was low (< 0.05) and the junction was sufficiently upregulated (fold change PSI-KD/PSI-control > 10).

Gene Ontology Analysis

Enrichment test for biological process, cellular component and molecular function terms was carried out as proposed by (Reimand et al., 2016), with G:Profiler and Enrichment Map Cytoscape plugin. We have compared genes identified by MAJIQ after filtering for probable and spurious events against a background of 14,594 multi-exon genes that were expressed (FPKM > 1) in any of the analyzed sample.

Calculation of MaxEnt Scores, ESE/ESS Composition, Putative Branch Points and Additional Analyses

We used the MaxEnt algorithm (Yeo and Burge, 2004) implemented in the Human Splicing Finder online tool (3.0) (Desmet et al., 2009) and maxentpy (https://github.com/kepbod/maxentpy) to calculate splice site strengths. Using the same tools, we analyzed the exonic splicing enhancer and silencer composition of the *RER1* reporter. Putative mammalian U2 branch points were predicted using the SVM-BPfinder online tool (Corvelo et al., 2010). The ASAP structure (PDB ID: 4A8X)(Murachelli et al., 2012) was visualized using PyMOL (1.8). The UpSet R package (Lex et al., 2014) and the Euler3 Applet (Rodgers et al., 2014) was used to visually compare the alternative splice analysis results. Sashimi-plots (Katz et al., 2015) were generated from data generated in this study, as well as available RNA-Seq (GEO accession number GSE63091) (Wang et al., 2014) data using the integrated function of the IGV (2.3.98) (Thorvaldsdóttir et al., 2013). The fitness scores were obtained from available data on HeLa cells generated by a high-resolution CRISPR-screen (Hart et al., 2015). The fitness score is a log Bayes factor (BF) for each gene, which was calculated with a Bayesian Analysis of Gene Essentiality algorithm (Hart et al., 2015). The confidence in the essentiality (higher impact on fitness) of the gene increases with more positive scores. The Bayes Factor cutoff at 5% false discovery rate for HeLa cells (15.47) was used to discriminate essential and non-essential genes.

Code Availability

For availability of codes that were developed for this project, please contact the corresponding authors.

QUANTIFICATION AND STATISTICAL ANALYSIS

p values < 0.05 were considered significant. Significance in all figures is indicated as follows: ns = not significant, *p = 0.01 to 0.05, **p = 0.001 to 0.01, ***p < 0.001.

RNP Immunoprecipitation

Signals of ^{32}P-labeled RNAs were scanned using a Typhoon FLA 7000 (GE Healthcare) and quantified using the ImageQuant TL software (GE Healthcare). Results are shown as mean ± SD. GraphPad Prism 5 was used to determine the statistical significance by one-way ANOVA with Bonferroni post-test.

End-Point RT-PCR

Bands detected in agarose gels from the indicated biological replicates of end-point PCRs were quantified using the Image Lab 6.0.1 software (Bio-Rad). Results of the indicated band % per lane are shown as mean ± SD. GraphPad Prism 5 was used to determine the statistical significance by one-way ANOVA with Bonferroni post-test. For experiments with only two samples, two-tailed unpaired Student's t test was performed with GraphPad Prism 5.

Quantitative RT-PCR

The reactions for each biological replicate were performed in duplicates and the average Ct (threshold cycle) value for retained or skipped *MRPL3* exon 4 was measured. Values for skipped exon were subtracted from values for retained exon to calculate the ΔCt. The fold changes were calculated using the ΔΔCt method, using the Luc knockdown as normalization (Schmittgen and Livak, 2008). The mean fold changes were calculated from three biologically independent experiments. Results are shown as mean ± SD. GraphPad Prism 5 was used to determine the statistical significance by one-way ANOVA with Bonferroni post-test.

Pathscan Assay

Chemiluminescent signals were measured with the ChemiDoc MP Imaging System (Bio-Rad) and quantified using the ImageQuant TL software (GE Healthcare). All signals were first normalized to the α-tubulin control signal. Then, mean values of three biological and two technical replicates, as well as differences between knockdown and control samples were calculated. Finally, the mean $\log_2$ of absolute relative signal intensities were plotted using the superheat R package (arXiv:1512.01524v2 [stat.AP]). Results are shown as mean ± SD. Propagation of error calculations were performed, and GraphPad Prism 5 was used to determine the statistical significance by one-way ANOVA with Bonferroni post-test.

Survival Assay

Background corrected mean luminescence was calculated and plotted using GraphPad Prism 5. Results are shown as mean ± SD. GraphPad Prism 5 was used to determine the statistical significance by two-way ANOVA with Bonferroni post-test.

DATA AND SOFTWARE AVAILABILITY

Data Resources

The accession number for the raw RNA-sequencing data reported in this paper is ArrayExpress: E-MTAB-6564.

Data Availability

The authors declare that all the data supporting the findings of this study are available within the article and its Supplementary Information files and from the corresponding authors upon reasonable request. The raw imaging data can be accessed via Mendeley: https://doi.org/10.17632/wt7ybwz82g.1.

Molecular Cell, Volume *72*

Supplemental Information

Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity

Volker Boehm, Thiago Britto-Borges, Anna-Lena Steckelberg, Kusum K. Singh, Jennifer V. Gerbracht, Elif Gueney, Lorea Blazquez, Janine Altmüller, Christoph Dieterich, and Niels H. Gehring

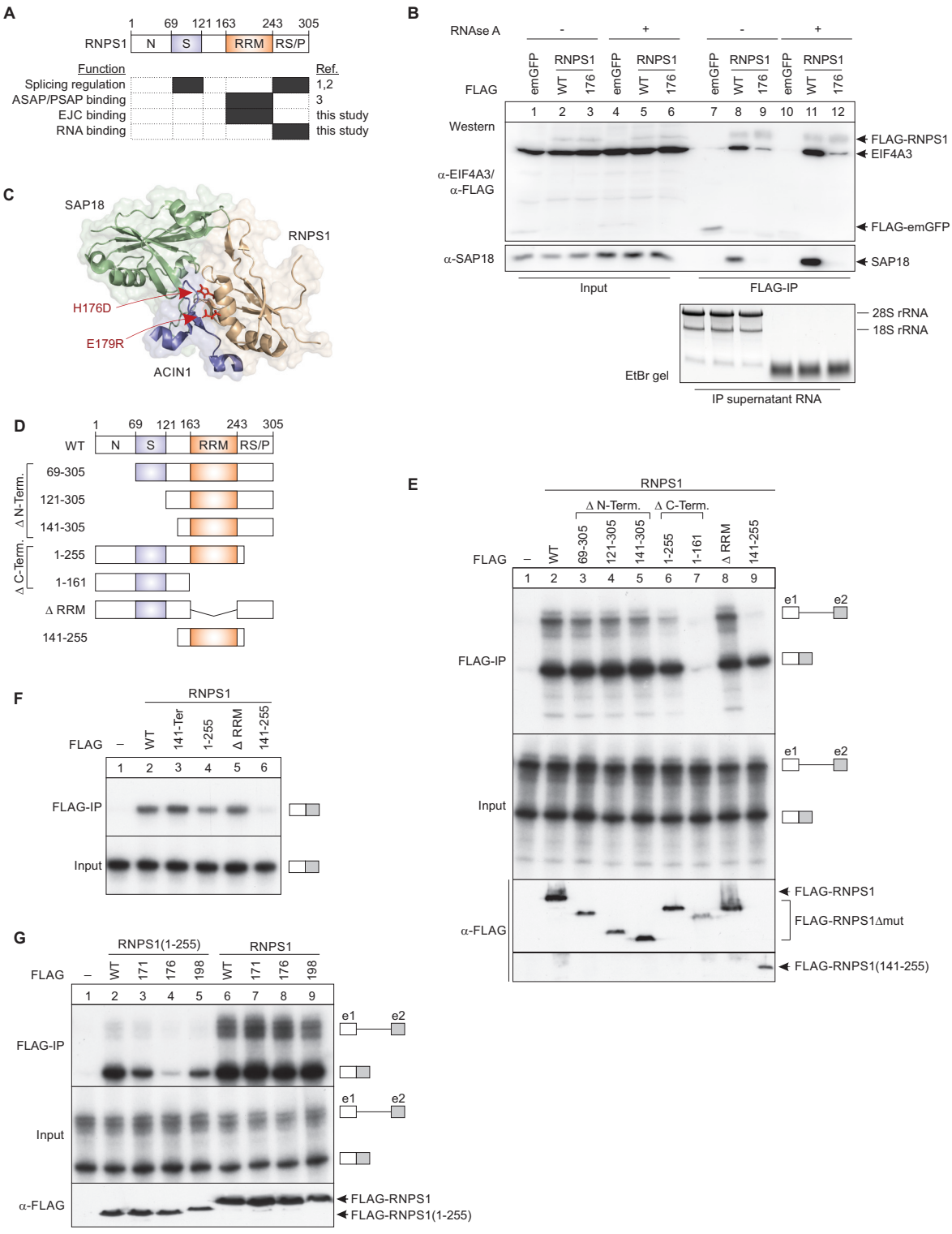


Figure S1. Characterization of RNPS1 binding to spliced mRNA, Related to Figure 1

(A) Schematic representation of RNPS1 domains and their functions. The indicated references are: 1 (Mayeda et al., 1999), 2 (Sakashita et al., 2004) and 3 (Murachelli et al., 2012).

(B) Co-immunoprecipitation of EJC core component EIF4A3 and ASAP/PSAP component SAP18 with the indicated FLAG-tagged RNPS1 variants in the presence or absence of RNAse A. n=1

(C) Location of the H176D-E179R mutation (red) on the surface of RNPS1 in the context of the ASAP complex components SAP18 and ACIN1.

(D), **(E)** In vitro splicing of 32P body-labeled MINX mRNA in the presence of FLAG-RNPS1 deletion variants **(D)**. After FLAG-immunoprecipitation of mRNPs, the RNA was extracted, resolved on a denaturing urea-gel and visualized by phosphorimaging **(E)**. Expression of the FLAG constructs was visualized by immunoblotting (bottom). n=1

(F) In vitro splicing of intron-less 32P body-labeled MINX mRNA in the presence of a subset of FLAG-RNPS1 deletion variants. n=1

(G) In vitro splicing of 32P body-labeled MINX mRNA in the presence of FLAG-RNPS1 WT or point mutants. n=1

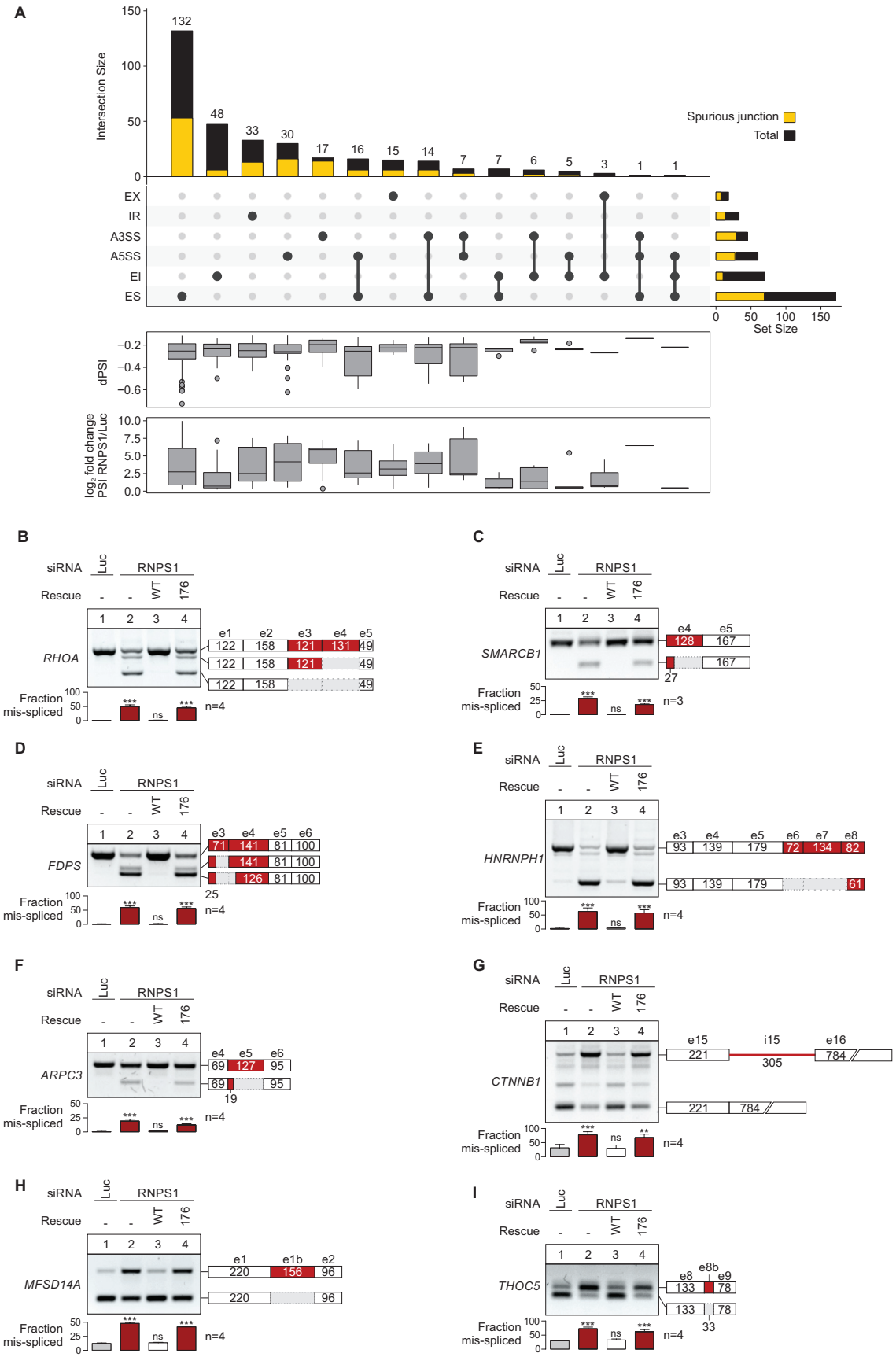


Figure S2. Validation of RNPS1-dependent alternative splice events, Related to Figure 2

(A) Complete intersection and classification of alternatively spliced junctions upon RNPS1 knockdown. The quantity of spurious junctions for each class is shown. The change in junction usage (delta percent spliced in; dPSI) and PSI fold change for each class is shown as boxplots on the bottom.

(B) - **(I)** RT-PCR analysis of the indicated splice events with RNA from HeLa cells expressing the indicated rescue proteins, transfected with the indicated siRNA. Exon-intron architecture for each target is depicted schematically, alternative sequences are highlighted. Quantified results from the indicated biological replicates are shown as mean ± SD and compared to the Luc knockdown control.

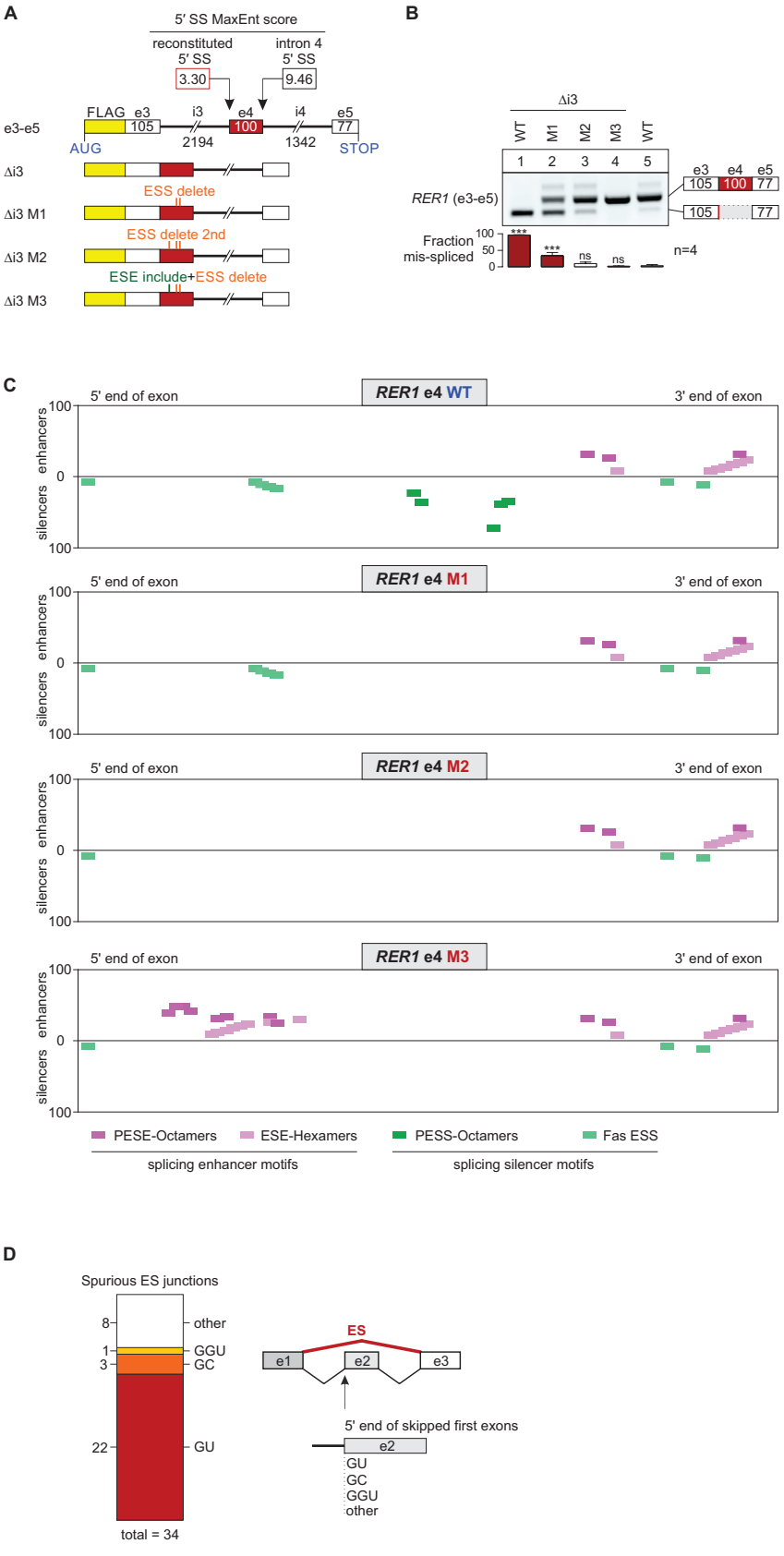


Figure S3. Characterization of the RER1 splice reporter, Related to Figure 3

(A) Overview of the *RER1* Δi3 reporter with exonic splice silencer (ESS) and exonic splice enhancer (ESE) mutations in e4.

(B) RT-PCR analysis of the *RER1* e4-mutated reporter constructs with RNA from HeLa cells. Quantified results from the indicated biological replicates are shown as mean ± SD and compared to the wildtype control.

(C) HSF analysis of ESS/ESE composition of the mutants depicted in **(A)**.

(D) Identification of 5′ terminal dinucleotides of first skipped exons in exclusive exon skipping events of spurious junctions.

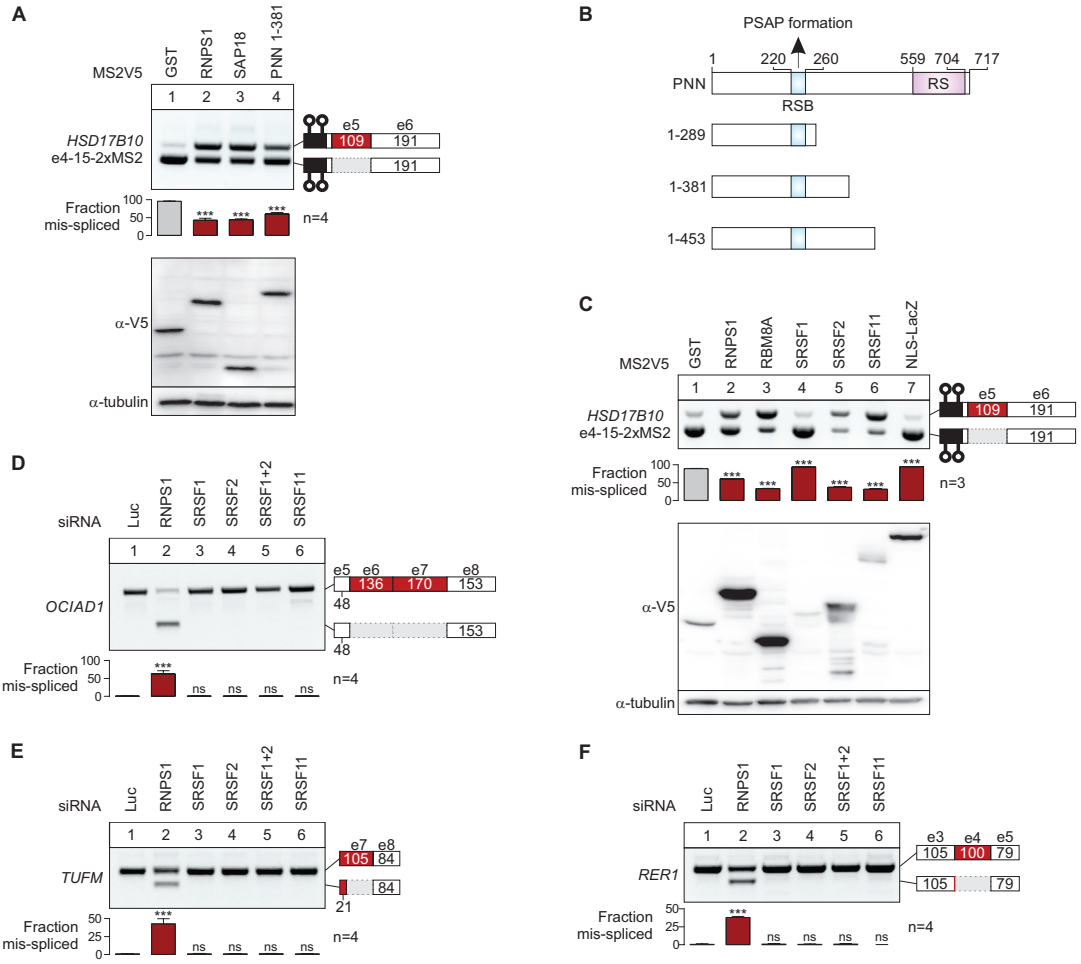


Figure S4. Insights into the mechanism of RNPS1- and PSAP-mediated splice site suppression, Related to Figure 4

(A) Reporter splicing from dual inducible stable HeLa cell lines expressing both the reporter *(HSD17B10* e4-15) and the indicated tethering protein was detected via RT-PCR.

(B) Scheme of PNN domain architecture and constructs used for tethering and rescue assays.

(C) The indicated MS2V5-tethering proteins and *HSD17B10* reporter were expressed transiently in HeLa Tet-Off cells and the splice patterns analyzed by RT-PCR.

(D) - **(F)** RT-PCR analysis of the indicated splice event with RNA from HeLa cells transfected with siRNA targeting various SR proteins.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

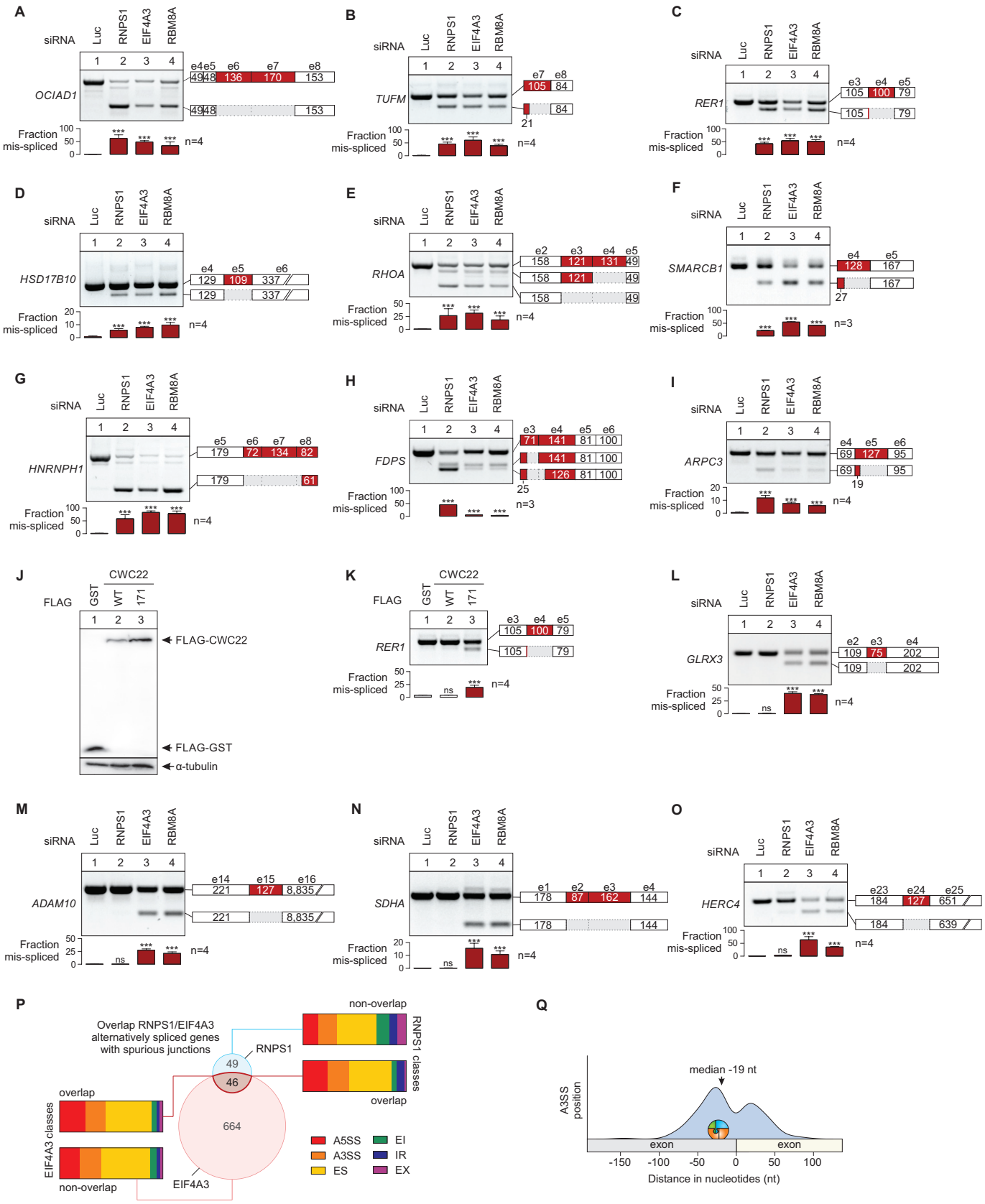


Figure S5. Validation of EJC-dependency of RNPS1-regulated splice events, related to Figure 5

(A) - **(I)** RT-PCR analysis of the indicated splice events with RNA from HeLa cells transfected with the indicated siRNA.

(J), **(K)** The effect of overexpression of CWC22 WT or NK171DE mutant in HeLa cells **(J)** on *RER1* alternative splicing was analyzed by RT-PCR **(K)**.

(L) - **(O)** RT-PCR analysis of the indicated splice events with RNA from HeLa cells transfected with the indicated siRNA.

(P) Overlap of genes with at least one spurious junctions identified in RNPS1 and EIF4A3 knockdown RNA-Seq data. The proportion of identified classes for the alternatively spliced genes are shown.

(Q) Depiction of EJC-dependent alternative 3′SS position of spurious junctions relative to exon boundaries as density plot. An outlier at -645 nt distance is not shown in this plot.

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

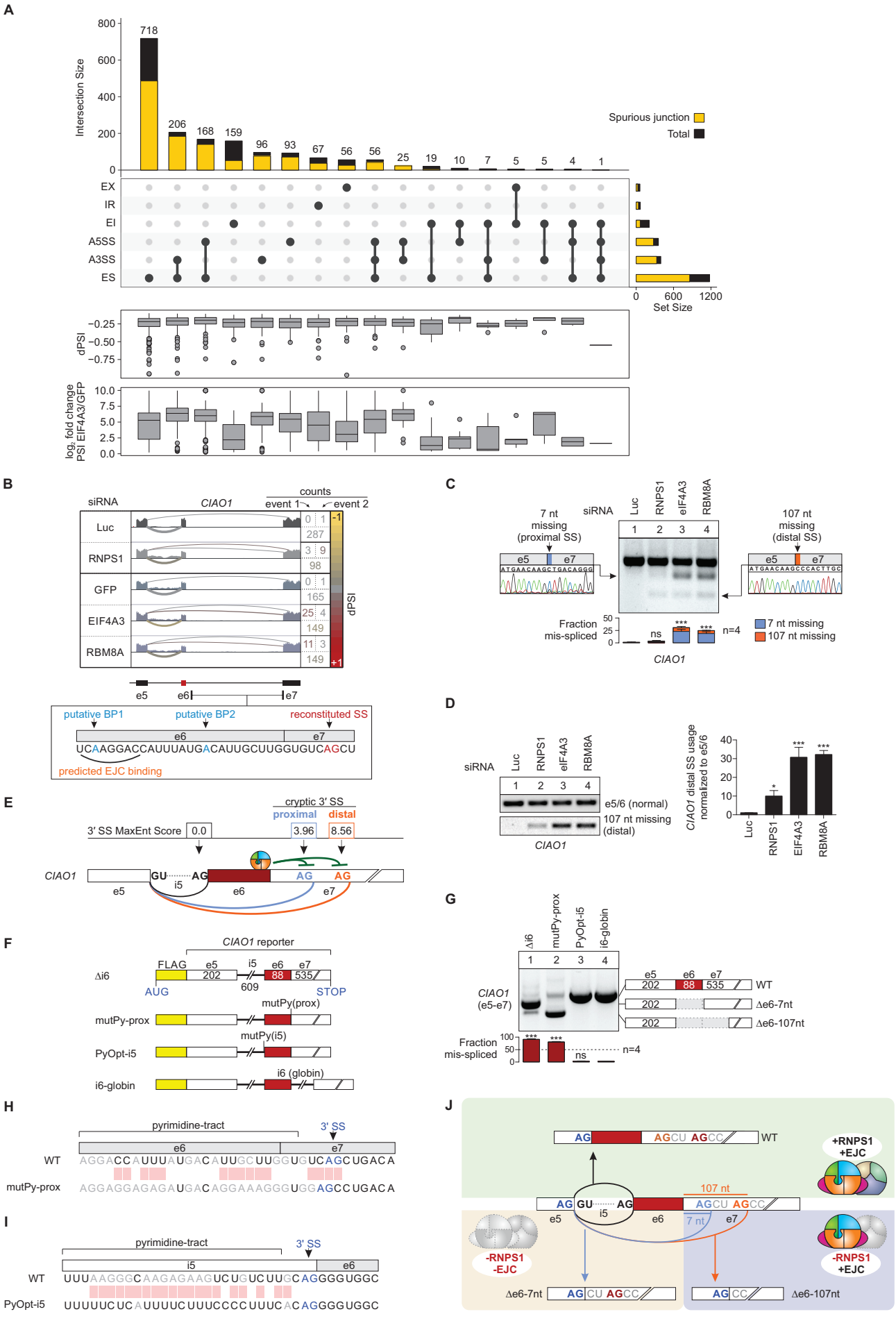

A
Intersection Size
Spurious junction
Total
EX
IR
EI
A5SS
A3SS
ES
Set Size
dPSI
log2 fold change PSI EIF4A3/GFP
B
siRNA
CIAO1
counts
event 1
event 2
Luc
RNPS1
GFP
EIF4A3
RBM8A
dPSI
putative BP1
putative BP2
reconstituted SS
predicted EJC binding
C
7 nt missing (proximal SS)
107 nt missing (distal SS)
Fraction mis-spliced
n=4
7 nt missing
107 nt missing
D
CIAO1 distal SS usage normalized to e5/6
e5/6 (normal)
107 nt missing (distal)
E
cryptic 3′ SS
proximal
distal
3′ SS MaxEnt Score
F
CIAO1 reporter
Δi6
mutPy-prox
PyOpt-i5
i6-globin
G
CIAO1 (e5-e7)
WT
Δe6-7nt
Δe6-107nt
H
pyrimidine-tract
3′ SS
I
J
+RNPS1 +EJC
-RNPS1 -EJC

Figure S6. Characterization of splice defects upon EJC-depletion, Related to Figure 5

(A) Complete intersection and classification of alternatively spliced junctions upon EIF4A3 knockdown. The quantity of spurious junctions for each class is shown. The change in junction usage (delta percent spliced in; dPSI) and PSI fold change for each class are shown as boxplots on the bottom.

(B) Sashimi-plots depicting two alternative splicing events for *CIAO1*. The predicted EJC binding site in relation to the cryptic splice site (SS) and putative branch points (BP) is indicated. Two different cryptic 3′SS in *CIAO1* exon 7 can be activated, resulting in splicing of exon 6 together with additional 107 (distal event) or 7 (proximal event) nucleotides from exon 7.

(C) *CIAO1* RT-PCR of cDNA samples obtained from HeLa cells transfected with the indicated siRNA. Sanger sequencing results of the individual PCR products are shown.

(D) RT-PCR of samples from **(C)** with *CIAO1* specific exon 5/6 or distal cryptic splice site primers. Distal cryptic splice site usage was normalized to exon 5/6 splicing in three biological replicates.

(E) Scheme of *CIAO1* canonical and cryptic splice sites highlighting the 3′ splice site scores. According to the hypothesis, the EJC deposited by intron 6 splicing suppresses the strong cryptic SS in exon 7 of *CIAO1* and thereby promotes the regular splicing of the suboptimal intron 5.

(F), **(G)** *CIAO1* minigene reporter constructs **(F) w**ere expressed in HeLa Tet-Off cells and analyzed via RT-PCR **(G)**. The *CIAO1* reporter lacking intron 6 (Δi6) was primarily spliced at the proximal cryptic SS. Mutating the polypyrimidine tract of the proximal cryptic SS (mutPy-prox; **(H)**) activated the distal cryptic 3′SS. Normal splicing was restored by enhancing the polypyrimidine tract of the genuine 3′SS of intron 5 (PyOpt-i5; **(I)**), or by inserting an intron downstream of exon 6 (i6-globin).

(H), **(I)** Mutated splice site consensus for the proximal cryptic splice site **(H)** and intron 5 canonical splice site **(I)**

(J) Scheme of *CIAO1* exon 5-7 intra-splicing in the presence of EJC without RNPS1 (bottom right) or absence of both RNPS1 and EJC (bottom left).

All data from the indicated biological replicates show the mean ± SD and were compared to the respective control.

Table S3. List of oligonucleotides used in this study, Related to STAR Methods

Oligo	Sequence	Purpose
MRPL3_e3_se	CCTTGAAGCTGGGCATGATG	RT-PCR
MRPL3_e6_as	GCAGCATAAAGAGGAGTGCC	
OCIAD1_e4_se	CCTTTGGCTGCAACAAGTATG	
OCIAD1_e9_as	TGCAGGTAATCATGACCACCT	
TUFM_e6_se	TTAGGCCTGAAGTCTGTGCA	
TUFM_e8_as	GCTTGATGGAACCTGGCTTG	
RER1_e2_se	GGGGACAGTGTGGGAGAATC	
RER1_e6_as	ATCACCAGAATCGGCCAGAA	
HSD17B_e4_se	TGGGCCAGAATGAACCAGA	
HSD17B_e6_as	CAGGGAAAGGAAGGGCAGA	
ACIN1_e13_se	CCAGGTGTCAGTAGAGGTGG	
ACIN1_e15_as	CCACCAAGGTTCCTGTGCG	
ATP5F1_e3_se	CCCTGTACCACCTCTTCCTG	
ATP5F1_e5_as	CCAGTGCCTGTTGTGACTTC	
ATP5B_e5_se	TGTTTGCTGGTGTTGGTGAG	
ATP5B_e7_as	GGATTCGGCCCAATAATGCA	
EIF2S3_e1_se	CGGGGTGATTTCCTTCCTCT	
EIF2S3_e12_as	TCCAACTTCCAAATCCATCCG	
DKC1_e1_se	GTTCCCTCGGCTGTGGAC	
DKC1_e15_as	CCAGCTTCAAGTGGCCTTC	
MPV17_e2_se	ACCCGTGGAAAGTACAGGTC	
MPV17_e8_as	ATGGAGTGAGGCAGGCTTAG	
ATL2_e1_se	GACGGACCAGCGACCCAA	
ATL2_e13_as	CTGTCGCTGTGCTGATGAAA	
RHOA_e1_se	TCCGTCGGTTCTCTCGTTAG	
RHOA_e5_as	CGCCAATCCTGTTTGCCATA	
SMARCB1_e3_se	ACCCTGTTAAAAGCCTCGGA	
SMARCB1_e5_as	CCCATCGATCTCCATGTCCA	
FDPS_e3_se	TCGTTAGGGTGCTGACTGAG	
FDPS_e6_as	TCAGGTAATAGGGCTGCTCC	
HNRNPH1_e3_se	CAAATAGTCCTGACACGGCC	
HNRNPH1_e8_as	ACTCGACATCTGCTTCACCA	
ARPC3_e3_se	GCCATCTATTACTTCAAGGCCA	
ARPC3_e6_as	TCTCAGTCCAGTCTCTTGCC	
CTNNB1_e15_se	AGGATGCCTTGGGTATGGAC	
CTNNB1_e16_as	TCTTGTGATCCATTCTTGTGCA	
MFSD14A_e1_se	GGAAGAAGAAGAAACGGGCC	
MFSD14A_e2_as	AGGTGGGTGCTGTCAATAGT	
THOC5_e8_se	CATTCGACCAGGCTCACAAG	
THOC5_e9_as	CTTCATCCACACTGCCTTCG	
GLRX3_e2_se	CGAAGTTATGGCAGAGTTAGCT	
GLRX3_e4_as	GCGTGGTTCTTGAGGAGTTC	
ADAM10_e14_se	GACCCATCAACTTGTGCCAG	
ADAM10_e16_as	TTGATAACTCTCTCGGGGCC	
SDHA_e1_se	CGGCAACAGCAGACATGTC	
SDHA_e4_as	TTGTCCTCCTCCATGTTCCC	
HERC4_e23_se	GCTTTTCATGCGGGCTTTCA	
HERC4_e25_as	TACTCCTCACCACCTCCTGT	
CIAO1_e5_se	TTGAGGGCCATGAATCCACT	
CIAO1_e7_as	CAGGCCACACAGTTGACATC	
CIAO1_e4-5_se	CACCCAAGTCAGGAGCTCTT	
CIAO1_e5-7_dist_as	GATGCAAGTGGGCTTGTTCATTG	
FLAG_se	ACAAGGACGACGATGACA	
BGH_as	TAGAAGGCACAGTCGAGG	

MRPL3_e3-e4-splice_se	TGGTCACATTACTTCAGGTACAA	qPCR
MRPL3_e3-e5-splice_se	CACATTACTTCAGAAAGCTACATCC	
MRPL3_e6_as	GCAGCATAAAGAGGAGTGCC	
VNN-(dT)20	TTTTTTTTTTTTTTTTTTTVNN	cDNA synthesis
TUFM_e6_Xho_se	TTTTCTCGAGTTAGGCCTGAAGTCTGTG	Reporter construction
TUFM_e9_Not_as	TTTTTTTGCGGCCGCTTACTCTGGGGGCAGGATAAT	
RER1_e3_rep_Xho_frame_se	TTTTTCTCGAGATTTATCAGTCCTGGCTAGACAAGTCC	
RER1_e5_rep_Sal_Not_stop_as	TTTTTTGCGGCCGCCCGGGTCGACTTAAAATTTAAACTCTGGGAGCCTTCGAATGA	
RER1_del_intron3_as	CATAGGTCACAATGTACCAACCCTGCAGCAGGTAAACTCGAA	
RER1_new_5SS_as	CAGTGCACAGCCGCTAACTTACCTGAGTCTTCCATTAAGGAAGG	
RER1_new_GT_as	CCCCAAGGCATAGGTCACTACCAACCAAGCCTGCAGCAG	
RER1_mutESS_as	GGATCCACTTTGGGAGAAAGGTAAGCTATGACGAGATTTAGATGGTAGATCCCCAG	
RER1_e4_M2-as	TATGACGAGATTTAGATGGTAGATCGCGAAGGCATAGGTCACAATGTACC	
RER1_e4_M3-ESE-as	AGCTATGACGAGATTTAGATGGTTGATCGTCTGCTTTCTGGTCACAATGTACCAACCCT	
HSD17B10_e2_rep_Xho_frame_se	TTTTTCTCGAGGGCCTGGTGGCGGTAATA	
HSD17B10_e6_rep_Sal_Not_as	TTTTTTGCGGCCGCCCGGGTCGACTATTAGGCACAGAGGGCGAC	
HSD17B10_del_intron4_se	GCTGCCTTCGAGGGTCAGGTTGGACAAGCTGCATACTC	
HSD17B10_e4-15_Xho_se	TTTTTCTCGAGGCCTTCGAGGGTCAGGTT	
2MS2_inert_HSD17B10_e4-30_as	GGCAGCCACACTGGCAGTCTCGACCGACGGCTGAT	
2MS2_inert_HSD17B10_e4-45_as	CAGTGTTGATGATGACCCCCTCGACCGACGGCTGAT	
ACIN1_e13_Xho_se	TTTTTCTCGAGGTAGTACCTGCAGAGGGCC	
ACIN1_e15_Sal_Not_as	TTTTTTGCGGCCGCCCGGGTCGACTTACGTTACAAAGCAATGAGATTTG	
ATP5F1_e3_Xho_se	TTTTTCTCGAGGTATTGCAGGCAACAAGGAC	
ATP5F1_del-i4_se	GACTTTGCTGATAAACTCAATGAGCAAAAACTTGCCCAACTAGAA	
ATP5F1_e5_Sal_Not_as	TTTTTTGCGGCCGCCCGGGTCGACTTACCTTTGCACATCAAAAAGGTA	
ATP5F1_e4_HA_se	TTTGTTGCAGACTTTGCTGATTACCCATACGATGTGCCCGATTACGCTAAACTCAATGAGGTAAGAACCATAA	
ATP5B_e5_Xho_se	TTTTTCTCGAGCTTTTTGGTGGTGCTGGA	
ATP5B_del-i6_se	ACCCAGGCTGGTTCAGAGGTGTCTGCATTATTGGGC	
ATP5B_e7_Sal_Not_as	TTTTTTGCGGCCGCCCGGGTCGACTTACTGTACAGAGGTGATAGATCCCTT	
ATP5B_e6_HA_se	CGCTTCACCCAGGCTGGTTACCCATACGATGTGCCCGATTACGCTTCAGAGGTAAGAGGGAAGGC	
CIAO1_e5_Xho_se	TTTTCTCGAGCTCTTAGCTTCTGCCAGCTATG	
CIAO1_i5e6_as	CCGCTGCATGCCACCCCTGCAAGAC	
CIAO1_i5_se	TTGAGGTGCCCAGGACATAG	
CIAO1_rep_Not_Sal_as	TTTTTTGCGGCCGCCCGGGTCGACTC	
CIAO1_new_Py-tract_se	CCCAGGACATAGGAACGTTTTTCTCATTTTCTTTCCCCTTTCACAGGGGTGGCATGCAGC	

CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex

Jennifer V. Gerbracht[1], Volker Boehm[1], Thiago Britto-Borges[2,3], Sebastian Kallabis[4], Janica L. Wiederstein[4], Simona Ciriello[1,5], Dominik U. Aschemeier[1], Marcus Krüger[4], Christian K. Frese[4,6], Janine Altmüller[7,8], Christoph Dieterich[2,3], Niels H. Gehring[1]

[1] Institute for Genetics, University of Cologne, 50674 Cologne, Germany
[2] Section of Bioinformatics and Systems Cardiology, Department of Internal Medicine III and Klaus Tschira Institute for Integrative Computational Cardiology, University of Heidelberg, 69120 Heidelberg, Germany
[3] DZHK (German Centre for Cardiovascular Research), Partner site Heidelberg/Mannheim, 69120 Heidelberg, Germany
[4] CECAD Research Center, University of Cologne, Joseph-Stelzmann-Str. 26, 50931 Cologne, Germany
[5] present address: AO Research Institute Davos, Clavadelerstrasse 8, CH-7270 Davos Platz, Switzerland
[6] present address: Max Planck Unit for the Science of Pathogens, 10117 Berlin, Germany
[7] Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, Germany
[8] Center for Molecular Medicine Cologne, University of Cologne, 50937 Cologne, Germany

Contact
Niels H. Gehring, University of Cologne, Institute for Genetics, Zuelpicher Str. 47a, 50674 Cologne, Germany; email: ngehring@uni-koeln.de

Running Title (40 Characters)
CASC3 promotes EJC-dependent NMD

Keywords (5, alphabetical order, separated by slash)
CRISPR-Cas9/gene expression/mRNA quality control/NMD/RNA degradation

Abstract

The exon junction complex (EJC) is an essential constituent and regulator of spliced messenger ribonucleoprotein particles (mRNPs) in metazoans. As a core component of the EJC, CASC3 was described to be pivotal for EJC-dependent nuclear and cytoplasmic processes. However, recent evidence suggests that CASC3 functions differently from other EJC core proteins. Here, we have established human CASC3 knockout cell lines to elucidate the cellular role of CASC3. In the knockout cells, overall EJC composition and EJC-dependent splicing are unchanged, whereas mRNA isoforms targeted by nonsense-mediated decay (NMD) are upregulated on a transcriptome-wide scale. Mechanistically, recruiting CASC3 to reporter mRNAs by direct tethering or via binding to the EJC stimulates mRNA decay and endonucleolytic cleavage at the termination codon. Building on existing EJC-NMD models, we propose that CASC3 equips the EJC with the ability to communicate with the NMD machinery in the cytoplasm. Collectively, our results characterize CASC3 as a peripheral EJC protein that tailors the transcriptome by promoting the degradation of EJC-dependent NMD substrates.

Introduction

Messenger RNA-binding proteins (mRBPs) determine the stability, location, and fate of bound mRNAs and are therefore important regulators of post-transcriptional gene expression (Hentze *et al.*, 2018). A central component of spliced mRNPs in metazoans is the exon-junction-complex (EJC), which is deposited during splicing upstream of exon-exon boundaries (Boehm & Gehring, 2016, Le Hir *et al.*, 2016, Woodward *et al.*, 2017). The heterotetrameric core of the EJC is composed of the proteins EIF4A3, MAGOH, RBM8A (Y14) and CASC3 (BTZ, MLN51) (Andersen *et al.*, 2006, Bono *et al.*, 2006). Generally, EJCs serve on spliced mRNAs as a mark that act as a binding platform for peripheral EJC-interacting factors (Singh *et al.*, 2012). The core and peripheral EJC components contribute to different steps of post-transcriptional gene expression including splicing regulation, mRNA localization, translation and nonsense-mediated mRNA decay (NMD) (Boehm & Gehring, 2016, Woodward *et al.*, 2017).

The EJC does not form spontaneously, but instead undergoes stepwise assembly in association with the spliceosome, while it proceeds through different spliceosomal complexes (Gerbracht & Gehring, 2018). As a first step, the splicing factor CWC22 recruits EIF4A3, to which the MAGOH/RBM8A heterodimer binds later on (Gehring *et al.*, 2009, Alexandrov *et al.*, 2012, Barbosa *et al.*, 2012, Steckelberg *et al.*, 2012, Steckelberg *et al.*, 2015). Unlike the three spliceosome-associated EJC components EIF4A3, MAGOH and RBM8A, the fourth protein CASC3 is absent from the purified spliceosomal C complex (Bessonov *et al.*, 2008). Furthermore, it cannot be detected in mRNPs formed on splicing intermediates (Gehring *et al.*, 2009). Therefore, it was suggested that CASC3 binds to the initially formed trimeric pre-EJC (consisting of EIF4A3, MAGOH and RBM8A) at a later stage. Interestingly, CASC3 has also been shown to be a shuttling protein that is mainly located in the cytoplasm, whereas the other EJC components are predominantly detected in the nucleus (Kataoka *et al.*, 2000, Le Hir *et al.*, 2001, Degot *et al.*, 2002, Palacios *et al.*, 2004, Shibuya *et al.*, 2004). It has therefore been suggested that CASC3 binds to the EJC in the nucleus and is transported with it into the cytoplasm (Ballut *et al.*, 2005). A recent study demonstrated that EJCs undergo a compositional switch

and that the ASAP/PSAP component RNPS1 and the protein CASC3 bind to functionally different mRNPs and exist in mutually exclusive EJCs (Mabin *et al.*, 2018). While EJCs of nuclear enriched transcripts were found to interact with RNPS1, the EJCs of cytoplasmic enriched transcripts rather contained CASC3. This observation is in line with the predominantly cytoplasmic localization of CASC3, but would argue against a nuclear function.

Another aspect under debate is the involvement of CASC3 in the NMD pathway. According to the EJC-dependent model of NMD, an EJC present more than 50-55 nucleotides downstream of a premature termination codon (PTC) triggers degradation by the NMD machinery (Nagy & Maquat, 1998). This quality control mechanism rids the cells from aberrant transcripts that contain PTCs due to mutations or mis-splicing. Additionally, it serves as a post-transcriptional mechanism of gene expression, especially when coupled to alternative splicing (Nasif *et al.*, 2018). This is a common feature of many genes coding for mRBPs, e.g. most SR proteins (Lareau *et al.*, 2007, Ni *et al.*, 2007). The EJC triggers NMD by interacting with members of the SURF complex resulting in phosphorylation of the central NMD factor UPF1 (Kashima *et al.*, 2006). Phosphorylated UPF1 then stimulates two distinct degradation pathways of NMD: The SMG5/7-dependent pathway results in deadenylation and decapping of the transcript followed by exonucleolytic decay from the 5’ end by XRN1 and the 3’ end by the exosome (Loh *et al.*, 2013). Alternatively, the transcript can be cleaved in the vicinity of the PTC by the endonuclease SMG6 which results in two mRNA fragments that can be exonucleolytically degraded by XRN1 and the exosome (Huntzinger *et al.*, 2008, Eberle *et al.*, 2009). While these pathways can act redundantly and in principle compensate for each other, SMG6-dependent endonucleolytic cleavage (endocleavage) has been shown to be the dominant pathways for NMD in human cells (Boehm *et al.*, 2014, Lykke-Andersen *et al.*, 2014, Colombo *et al.*, 2017). In cells depleted of CASC3 a stabilizing effect on PTC-containing reporter mRNAs and selected endogenous targets was reported (Palacios *et al.*, 2004, Gehring *et al.*, 2005, Gehring *et al.*, 2009). Furthermore, tethering CASC3 to an mRNA results in UPF1-dependent degradation of the transcript (Gehring *et al.*, 2009). However, a

recent report has challenged these observations and showed that CASC3 plays a minor role in NMD and only for certain endogenous targets in contrast to EIF4A3 or RNPS1 (Mabin *et al.*, 2018).

We were intrigued by the contrasting reports about the enigmatic role of CASC3 and decided to investigate the function of CASC3 and its distinction to the other EJC core components in more detail. For this purpose, we established HEK 293 CASC3 knockout (KO) cells using CRISPR-Cas9-mediated gene editing. The CASC3 KO cell lines are unchanged in their composition of the EJC core and peripheral interacting proteins. However, RNA-sequencing reveals an upregulation of transcript variants containing premature-termination codons (PTC) as well as the differential expression of many known NMD targets, indicating a perturbation of this decay pathway. Mechanistically, CASC3 stimulates SMG6-dependent turnover of NMD targets and likely acts as a link from the EJC to the NMD machinery. On the basis of these results we propose a revised model of EJC-dependent NMD in human cells.

Results

CASC3 is dispensable for nuclear EJC functions

Previously, we and others have shown that depletion of EJC core components in human cells leads to pervasive re-splicing of cryptic splice sites, resulting in aberrant splice variants lacking exonic sequences (schematically depicted in Supplementary Figure 1) (Wang *et al.*, 2014, Boehm *et al.*, 2018). Mechanistically, the EJC prevents the use of cryptic splice sites either by interaction with the ASAP/PSAP component RNPS1 or by sterically masking splice sites (Boehm *et al.*, 2018). In keeping with previous observations from HeLa cells, the knockdowns of the EJC core components EIF4A3 or RBM8A in HEK293 cells resulted in exon-skipping of the mRNAs for RER1, OCIAD1 and MRPL3 (Figure 1 A-C). Surprisingly, a knockdown of the EJC core factor CASC3 did not result in mis-splicing of these three selected transcripts. (Figure 1A-C). However, the knockdown of CASC3 increased the abundance of an alternative transcript isoform of the SR protein SRSF2, which is known to be targeted by NMD (Figure 1D) (Sureau *et al.*, 2001, Wollerton *et al.*, 2004). Together, these results indicate that CASC3 contributes to NMD activation by the EJC, but is not required for EJC-dependent splicing regulation. Given that the generation of EJC-specific splice isoforms is a readout for EJC assembly, the initial formation of EJCs does not seem to be affected by the knockdown of CASC3.

EJC assembly is unaffected in CASC3 knockout cells

Although the knockdown efficiency of CASC3 was substantial (Figure 1E), we wished to exclude that residual amounts of CASC3 prevented a reliable assessment of the protein's function. Since CASC3 was found to be non-essential in multiple genome-wide screens of human immortalized cell lines, we reasoned that knockout (KO) of CASC3 should be feasible (Hart *et al.*, 2015, Hart *et al.*, 2017). Accordingly, we obtained three cell lines by CRISPR-Cas9-mediated gene editing, designated H, F, and T lacking the CASC3-specific 130 kDa band on a western blot (Figure 2A). In all cell lines we detected genomic insertions of different length and sequence at the

beginning of the coding region of CASC3, which resulted in frame shifts of the downstream coding region or, in the case of cell line T, contained in-frame termination codons (Figure 2B, Supplemental Figure 2A and B). For the cell lines H and F, we observed an additional band of 100 kDa on western blots with antibodies recognizing the C-terminal or central region of CASC3 (Figure 2A, red arrow, Supplementary Figure 2A and C). This cross-reactive protein interacted with FLAG-tagged EIF4A3 and disappeared upon treatment with siRNAs against CASC3 (Figure 2C, Supplementary Figure 2C and D). This suggests that the cell lines H and F produce an N-terminally truncated form of CASC3, presumably representing a novel, non-canonical protein product. Such aberrant proteins have been recently described in a systematic analysis to be common by-products of CRISPR-Cas9 genome editing (Tuladhar *et al.*, 2019).

To assess the composition of exon junction complexes and their peripheral interacting proteins in the absence of CASC3 KO, we analyzed the FLAG-tagged EIF4A3 interactome in the cell line H and wild type cells using mass spectrometry. EIF4A3 was successfully enriched, together with other known EJC complex members (Supplemental Figure 2E). Although co-precipitated CASC3 was strongly reduced in the knockout cell line (Supplemental Figure 2G, log2 fold change = -3.75), it was enriched relative to the control (Supplemental Figure 2F, log2 fold change = 2.43). This confirmed that residual amounts of CASC3 protein are still present in the knockout cell line H. When the cells were treated additionally with CASC3 siRNAs, the amount of CASC3 pulled down by EIF4A3 was reduced to background levels (Figure 2D and E). Therefore, the cell lines H and F are comparable to a constitutive knockdown of CASC3, whereas the cell line T (without further treatment) and cell line H in combination with CASC3 siRNA treatment completely lack detectable CASC3 protein. This set of cell lines enables a hitherto unfeasible analysis of CASC3's cellular function as part of the EJC.

Strikingly, no other EJC core factor or splicing regulatory EJC component (e.g. ASAP/PSAP) was significantly altered in the CASC3-depleted condition (Figure 2F). This is in line with our initial conclusion that splicing-regulatory EJCs can assembly even in the absence of CASC3. Further support comes from co-immunoprecipitation analysis of EIF4A3, which demonstrated unchanged interaction with the EJC component MAGOH in the CASC3 knockout cell line H (Supplementary Figure 2H). In agreement with the data obtained from the knockdown of CASC3 in Figure 1, transcripts containing EJC-dependent splice sites were correctly spliced in CASC3 KO cells (Figure 2G and H, Supplementary Figure 2I and J), indicating that CASC3 is not necessary for correct and stable deposition of the EJC during splicing and its subsequent nuclear splice function. The role of CASC3 during EJC assembly was also characterized by performing *in vitro* splicing of a MINX intron in cell lysates obtained from the cell line H (Figure 2I and Supplementary Figure 2K). Notably, the overall splicing efficiency of the MINX transcript was slightly reduced in the input of CASC3 KO lysates when compared to wild-type lysates. However, spliced RNA was pulled down to the same extent by EIF4A3 (2.97% in KO vs. 3.01% in WT), suggesting that EIF4A3 is deposited on spliced mRNA even when CASC3 is strongly depleted. In the absence of CASC3, the only other protein that was significantly, although mildly reduced in EIF4A3-IPs was the NMD factor UPF3B (Figure 2F, log2 fold change = -1.52). UPF3B links the EJC to the NMD machinery via direct interactions (Buchwald *et al.*, 2010) and was recently found to be enriched in cytoplasmic CASC3-loaded EJCs (Mabin *et al.*, 2018). The reduction of NMD-competent EJCs could contribute to the NMD impairment that we observed upon loss of CASC3.

CASC3 regulates NMD-sensitive isoforms

To investigate the global effects of CASC3 depletion on the transcriptome, we performed RNA-sequencing (RNA-seq) of the cell lines H and T either treated with CASC3 or control siRNAs (Figure 3A) and identified differentially expressed genes (Supplementary Figure 3A and 3B, Supplementary Table 4). To exclude clone-specific and siRNA treatment-related effects, we compared the identified targets between the four conditions. Overall, the high number and the substantial

184 overlap of upregulated genes suggests that CASC3 KO mainly results in the accumulation of certain
185 transcripts (Figure 3B and C, Supplementary Figure 3C). This observation fits with the hypothesis
186 that CASC3 is required for efficient execution of NMD. Accordingly, several upregulated genes
187 belong to the class of small RNA (e.g. snoRNA) host genes, which are frequently NMD targets
188 (Figure 3C and D) (Lykke-Andersen *et al.*, 2014). We validated the upregulation of the snoRNA host
189 gene ZFAS1 by qPCR, which was even more pronounced in CASC3 KO than UPF1 knockdown cells
190 (Supplementary Figure 3D and E). Across the top 100 significantly upregulated genes in CASC3 KO,
191 15 small RNA host genes were identified (Figure 3D). Comparing the differentially expressed genes
192 to recent transcriptome-wide NMD screens, many of the top 100 significantly upregulated genes
193 were also differentially expressed in UPF1 and SMG6/7 KD (Figure 3D, 23% and 59%, respectively)
194 (Colombo *et al.*, 2017, Data ref: Colombo *et al.*, 2017). However, none of the identified targets
195 were present in an RNPS1 knockdown (Figure 3D) (Boehm *et al.*, 2018, Data ref: Boehm *et al.*,
196 2018). Collectively, our differential gene expression analysis strengthens the proposed link
197 between CASC3 and the NMD-machinery.

198 Next, we analyzed alternative splicing changes in CASC3 KO cells (Figure 3A). Since our earlier
199 assays showed that CASC3 was not involved in EJC-regulated splicing (Figures 1 and 2), we were
200 surprised to detect many altered splicing events in CASC3 depleted cells (Supplementary Figure 3F,
201 Supplementary Table 5). It is remarkable that hardly any alternative splicing events were shared
202 between RNPS1 knockdown and CASC3 KO cells (Figure 3E, Supplementary Figure 3F). Either
203 CASC3 regulates an RNPS1-independent set of alternative splice sites or the splicing changes are
204 due to impaired NMD, which fails to remove NMD-sensitive isoforms. To test these possibilities,
205 we investigated the functional consequence of CASC3-dependent alternatively splicing on the
206 transcript isoform level (Figure 3A, Supplementary Figure 4A and B, and Supplementary Table 6).

207 Strikingly, in all CASC3 KO conditions the majority of upregulated mRNA isoforms contained a
208 premature termination codon (PTC), rendering the transcripts susceptible to NMD (Figure 4A). On

the other hand, downregulated isoforms rarely contained a PTC. Among the identified isoform switches was the NMD target SRSF2, which we confirmed earlier to be CASC3-dependent (Figure 1D). While overall SRSF2 gene expression varied only slightly between wild-type and CASC3 KO cells, the isoform usage changed dramatically towards the accumulation of NMD-sensitive transcripts in the CASC3 KO conditions (Figure 4B).

We next validated a set of transcript isoform switches by qPCR (Figure 4C-E, Supplementary Figure C and D). In the transcript isoforms stabilized by the CASC3 KO and by a UPF1 KD, the inclusion of intronic regions resulted in the inclusion of PTCs. For example, an NMD-sensitive transcript variant of the gene TOE1 is produced alongside the canonical isoform by usage of an alternative 3' splice site (Supplementary Figure 4D). This event was also prominently detected in the SMG6/7 and UPF1 dataset (Supplementary Figure 4E).

The shift of isoform usage from NMD-insensitive to PTC-containing transcripts was also observed transcriptome-wide in CASC3 KO cells and was comparable to NMD-compromised SMG6/7 or UPF1 depleted cells (Figure 4F). These findings indicate that many transcript isoforms upregulated upon depletion of CASC3 represent genuine endogenous NMD targets. On the gene level, nearly 50% of CASC3-dependent alternatively used transcript isoforms were also found in the SMG6/7 or UPF1 conditions (Figure 4G and Supplementary Figure 4F). The incomplete overlap between CASC3 and SMG6/7 or UPF1 could be due to technical differences of the data sets or because not all NMD sensitive transcripts in the cell are affected by a knockout of CASC3. This could include transcripts that are degraded by EJC-independent NMD (Buhler *et al.*, 2006, Metze *et al.*, 2013).

CASC3 stimulates SMG6-dependent endocleavage

To deepen the understanding of how a lack of CASC3 results in reduced NMD efficiency, we stably integrated the well-established globin NMD reporter PTC39 in WT and CASC3 KO cell lines (Figure 5A and B, Supplementary Figure 5A) (Thermann *et al.*, 1998). The analysis of a reporter mRNA enables a read-out of multiple aspects of mRNA degradation: firstly, the total levels of the

full-length reporter, secondly the contribution of 5'->3' exonucleolytic decay by XRN1 (detection of xrFrag due to an XRN1-resistant element (Boehm *et al.*, 2016, Voigt *et al.*, 2019); and thirdly the amount of endonucleolytic cleavage by SMG6 (detection of 3' fragment stabilized by XRN1 knockdown). In both WT and CASC3 KO cell lines the reporter was efficiently degraded, showing that the NMD pathway is still functional in CASC3 depleted cells. However, full-length reporter levels in CASC3 KO were slightly higher when compared to wild-type cells (lane 2 vs. lane 5). Notably, the accumulation of 3' fragments following XRN1 knockdown was clearly reduced in the CASC3 KO condition (lane 3 vs. lane 6). This difference in endocleavage efficiency was also observed when expressing a minigene reporter of the endogenous CASC3 target TOE1. While there was a substantial upregulation of full-length reporter mRNA in CASC3 KO cells, the amount of the 3' fragment was strongly reduced, suggesting that SMG6-dependent endocleavage is inefficient in CASC3 KO cells (Figure 5C and D, Supplementary Figure 5B-D). To further address this, a TPI reporter was expressed in combination with knockdowns of XRN1, SMG6 and/or SMG7 (Figure 5E-G). The degree of reporter and 3' fragment stabilization of the TPI reporter following XRN1 knockdown was comparable to the observations made for the globin reporter (lanes 1-3 vs. lanes 6-8). In both WT and CASC3 KO cells, SMG6 knockdown resulted in a drastic reduction of 3' fragments, as expected (lanes 4 and 9). Notably, a knockdown of SMG7 together with XRN1 revealed a major difference between the cell lines. In WT cells the PTC-containing reporter was only minimally stabilized by the SMG7/XRN1 knockdown and 3' fragments were unaffected. Performing a SMG6/XRN1 knockdown in CASC3 KO cells lead to a more dramatic stabilization of the full-length reporter and a decrease of 3' fragments compared to the XRN1 knockdown condition (lanes 3 and 5 vs. lanes 8 and 10). Collectively, our results indicate that in CASC3 KO cells SMG6-mediated endocleavage is impaired. This could explain why the CASC3 KO cells are more sensitive to a knockdown of SMG7 when compared to wild type cells.

To identify, which part of CASC3 promotes NMD, we employed a tethering reporter that was designed to monitor mRNA turnover as well as endocleavage at the termination codon (Figure 6A).

Tethering the full-length CASC3 protein to the MS2 stem loops downstream of the stop codon resulted in degradation of the reporter compared to tethering of the negative control GST (Figure 6B and C, Supplementary Figure 6A). This degradation was accompanied by the production of 3' fragments in XRN1 knockdown conditions, indicating that the mechanism of decay is comparable to the PTC-containing reporter mRNAs (Figure 6B, lane 6). Surprisingly, C-terminally truncated deletion mutants of CASC3 that contain the first 480 or even 137 amino acid residues were able to induce degradation of the tethering reporter to a comparable extent as full-length CASC3 (Figure 6B, lanes 3 and 4, 7 and 8).

Finally, CASC3 deletion mutants were expressed in the CASC3 KO cells to identify the minimal part necessary to rescue the effects on endogenous NMD targets (Figure 6D, Supplementary Figure 6B and C). As in the tethering experiment, expression of full-length CASC3 and the C-terminal truncated variant 1-480 resulted in transcript isoform levels comparable to wild-type cells for the targets CLN6 and TOE1 (Figure 6D lanes 1-4). An EJC binding-deficient mutant of CASC3 (188/218 double point mutation) was unable to rescue, supporting the notion that CASC3 is recruited to the mRNA by binding to the EJC (Figure 6D, lane 5). Deleting the N-terminal 109 amino acids of CASC3 (110-480) did not alter the rescue ability (Figure 6D, lane 6). While in the tethering assay it was sufficient to place the N-terminus downstream of a termination codon, this part of CASC3 was not necessary to rescue NMD activity in the KO cells. This suggests that different domains of CASC3 act in a redundant manner during the activation of NMD by the EJC.

Discussion

The role of CASC3 within the EJC has been the subject of scientific controversy for many years. CASC3 has been initially described as an EJC core protein, because it was required for the assembly of the EJC from recombinant protein components *in vitro* (Ballut *et al.*, 2005). However, it has been demonstrated that the mechanism of EJC assembly using recombinant proteins is mechanistically different from EJC assembly in splicing extracts or in living cell (Steckelberg *et al.*, 2012). Furthermore, several recent publications challenged the view of CASC3 being an EJC core component. For instance, CASC3 was reported to be present in substoichiometric amounts compared to the other three EJC core proteins EIF4A3, RBM8A, and MAGOH in HEK293 (Singh *et al.*, 2012) and U2OS cells (Beck *et al.*, 2011). Also, during mouse embryonic brain development CASC3 deficiency results in a different phenotype than the other EJC core components (Mao *et al.*, 2017). By using CASC3 CRISPR-Cas9 knockout cells, we unambiguously establish that CASC3 is not required for EJC assembly or EJC-regulated splicing in the nucleus (Figure 7). Therefore, our molecular analyses fully support the recently emerging view of defining CASC3 as a peripheral EJC component. As a mainly cytoplasmic component of the EJC we propose that the principal role of CASC3 is to alter the efficiency by which NMD-sensitive transcript isoforms are degraded.

Although NMD has been extensively studied in the past decades and many NMD factors have been identified and characterized, no universal model exists that describes how they work together to elicit NMD. While a function of CASC3 in NMD has been reported before, previous analyses did not show consistent results, ranging from a substantial contribution of CASC3 to only a minor role in NMD (Palacios *et al.*, 2004, Gehring *et al.*, 2005, Gehring *et al.*, 2009, Mabin *et al.*, 2018). In addition, none of the previous publications performed transcriptome-wide analyses but concentrated on reporter mRNAs of only a few selected endogenous NMD targets. We reasoned that the inconsistent results in the literature may be influenced by variable CASC3 knockdown efficiency. By generating CASC3 knockout cell lines, we can for the first time analyze the global effects of a complete depletion of CASC3 on the transcriptome and can exclude that residual CASC3 protein masks these effects. Interestingly,

the CASC3 knockout had no deleterious effect on the viability of HEK293 cells, unlike the depletion of the EJC components EIF4A3, RBM8A, or MAGOH. Nonetheless, since CASC3 is required for mouse embryogenesis and involved in the transport of mRNAs in *D. melanogaster*, it is likely that CASC3 downregulation in highly specialized cell types such as neurons or in developing tissues would result in a more severe phenotype compared to HEK293 cells.

In recent years, high-throughput RNA-sequencing became an increasingly important method for the analysis of NMD. Several RNA-Seq datasets of cells with NMD-factor knockdowns have been generated and analyzed (Tani *et al.*, 2012, Hurt *et al.*, 2013, Lykke-Andersen *et al.*, 2014, Schmidt *et al.*, 2015, Colombo *et al.*, 2017). However, these datasets were obtained in different cell lines, with different amounts of replicates and due to the rapid developments of next-generation sequencing, not using the same technologies. Furthermore, batch effects and divergent approaches of data analyses may contribute to the fact that only a minor overlap of NMD targets could be established so far (Colombo *et al.*, 2017). We compared the results of our CASC3 KO RNA-sequencings to the most recent and comprehensive NMD factor analysis performed by Colombo *et al.* (2017). The differential expression analysis revealed that many of the top upregulated genes in the CASC3 KO datasets are also significantly affected by UPF1 or SMG6/7 knockdowns (Colombo *et al.*, 2017) and/or encode for small RNA (sRNA) host genes, a previously described class of NMD targets (Lykke-Andersen *et al.*, 2014).

We detected many alternative splicing events in the CASC3 KO data, which was unexpected given that CASC3 was apparently dispensable for the nuclear EJC-related functions. However, we could attribute these splicing patterns to dysfunctional NMD, since isoform-specific algorithms revealed that predominantly PTC-containing transcripts accumulated. Using all the available bioinformatics analysis pipelines and a systematic approach to detect affected transcripts under NMD factor knockdown/knockout conditions could therefore be a crucial step to paint a complete picture of the regulation of transcripts by NMD in the future.

How exactly CASC3 activates NMD when bound to an EJC, is not yet fully understood. Previously, we reported that the presence of EJCs in the 3' UTR enhances endocleavage (Boehm *et al.*, 2014). In line with the proposed role as a peripheral NMD-activating EJC component, we observed that CASC3 stimulates SMG6-dependent endocleavage, thereby promoting the degradation of NMD-targeted transcripts. This effect can be recapitulated by tethering full-length CASC3, its N-terminal two thirds (1-480) or just its N-terminal 137 amino acids to a reporter mRNA. How the small N-terminal region, which cannot assemble into the EJC or contains any known protein domains or sequence motifs can elicit NMD remains to be determined. It is also unclear, if the N-terminus activates translation-dependent degradation, as it was previously shown for the full length CASC3 (Boehm *et al.*, 2016). Since the N-terminal segment of CASC3 is a region of low-complexity it could hypothetically undergo liquid-liquid phase separation (LLPS) and be present in condensates with mRNA decay factors, such as processing bodies (P-bodies). In agreement with this idea CASC3 was shown to localize to cytoplasmic granules when overexpressed (Cougot *et al.*, 2014).

Our data suggest that CASC3 activates NMD by potentially redundant mechanisms. This fits very well into the general picture of NMD, which uses multiplexed degradation pathways to efficiently and robustly remove mRNAs. Mechanistically, binding of CASC3 to the EJC could have an indirect effect on NMD stimulation by increasing the stability of the bound EJC and thus maintaining the possibility of efficient endonucleolytic cleavage of the transcript. An indication for this role comes from the initial *in vitro* observation that CASC3 stabilizes recombinant EJCs (Ballut *et al.*, 2005). Additionally, the moderately reduced pull-down of UPF3B with EIF4A3 could indicate that cytoplasmic NMD-competent EJCs are less stable in CASC3 KO cells. Alternatively, CASC3 may directly contribute to the recruitment of NMD factors. We therefore propose that CASC3 either alone or in conjunction with UPF3B links the EJC with the NMD machinery. In particular, CASC3 influences the contribution of SMG6-mediated endonucleolytic and SMG7-dependent exonucleolytic decay pathways to the overall degradation efficiency of NMD. Accordingly, in wild type cells a knockdown of SMG7 only had a marginal effect on the abundance of the analyzed NMD reporter mRNA, whereas it clearly impaired NMD in CASC3 KO

355 cells. Also, the amount of endocleavage-derived 3' fragments was reduced when CASC3 is depleted,
356 mirroring the SMG6-knockdown condition.

357 By integrating CASC3 as a specific NMD-activating factor we can now postulate a modified model
358 of EJC-dependent NMD, which is also compatible with several molecular properties of the EJC (Figure
359 7). Since CASC3 is only present in modest amounts in the cytoplasm, it will probably not immediately
360 associate with all EJCs on recently exported mRNPs. This would also not be necessary, since most EJCs
361 are located in the coding sequence and will therefore be removed by the first translating ribosome.
362 However, mRNAs containing PTCs will carry one or more EJCs in their 3' UTR, which are available for
363 binding of CASC3. The first few translating ribosomes may terminate upstream of CASC3-free EJCs,
364 which could preferentially trigger SMG7-dependent exonucleolytic degradation. Previously, NMD has
365 been proposed to occur primarily in the pioneering round of translation when newly synthesized
366 transcripts are bound to the cap-binding complex (Ishigaki *et al.*, 2001, Maquat *et al.*, 2010). This model
367 has been challenged and there is evidence that NMD can occur on already translating mRNAs and
368 possibly with a constant probability during every round of translation (Durand & Lykke-Andersen,
369 2013, Rufener & Muhlemann, 2013, Hoek *et al.*, 2019). Thus, CASC3 could bind to the EJC at a later
370 time point and then increase the probability to activate SMG6-mediated endonucleolytic degradation
371 after each round of termination. Important molecular targets of CASC3 may be mRNAs that escape
372 initial NMD activation, despite containing a PTC (Trcek *et al.*, 2013, Hoek *et al.*, 2019). CASC3 may help
373 to reduce the amount of these mRNAs by maintaining the NMD-activating function of the EJC, either
374 by increasing its stability on the mRNA or via direct interactions with the NMD machinery. This concept
375 would be consistent with the recent observation that NMD targets undergo several rounds of
376 translation before endocleavage occurs (Hoek *et al.*, 2019).

377 In summary, our data paint a picture, in which CASC3 has no essential EJC-related function in the
378 nucleus, but helps to turn EJCs into a persistent RNA degradation mode. We do not exclude the
379 possibility that CASC3 is already associated with the EJC in the nucleus. However, our model of delayed

380 binding of CASC3 to the EJC in the cytoplasm would explain why only a small amount of CASC3 is
381 sufficient to activate EJC-dependent NMD. In this model, CASC3 is an indispensable cytoplasmic
382 component of the EJC that helps to degrade mRNAs that failed to unload all their bound EJCs during
383 the initial rounds of translation. Thus, the binding of CASC3 to the EJC could signal the final round(s) of
384 translation of an mRNA.

Materials and Methods

Cell culture

Flp-In 293 T-REx cells (Thermo Fisher Scientific) were maintained at 37° C, 5% CO_2 and 90% humidity in Dublecco's Modified Eagle Medium (DMEM, Thermo Fisher Scientific) supplemented with 9% fetal bovine serum (FBS) and Penicillin-Streptomycin (both Thermo Fisher Scientific). Tethering experiments were performed in HeLa Tet-Off cells (Clontech) cultured in the same conditions.

siRNA-mediated knockdowns

The cells were seeded in 6-well plates at a density of $2x10^5$ cells per well and reverse transfected using 2.5 µl Lipofectamine RNAiMAX and 60 pmol of the respective siRNA(s) according to the manufacturer's instructions. In preparation for mass spectrometry, the cells were reverse transfected in 10 cm dishes using 10 µl Lipofectamine RNAiMAX and 300 pmol siRNA. siRNAs were targeted against Luciferase (5'-CGTACGCGGAATACTTCGA-3'), EIF4A3 (5'-AGACATGACTAAAGTGGAA-3'), RBM8A (5'-TTCGCAGAATATGGGGAAA-3'), CASC3 (5'-CTGATGACATCAAACCTCGAAGAAT-3', 5'-CGTCATGAACTTTGGTAATCCCAGT-3'), UPF1 (5'-GATGCAGTTCCGCTCCATT-3'), XRN1 (5'-AGATGAACTTACCGTAGAA-3'), SMG6 (5'-GGGTCACAGTGCTGAAGTA-3') or SMG7 (5'-CGATTTGGAATACGCTTTA-3').

Generation of knockout cells using CRISPR-Cas9

The knockouts were performed using the Alt-R CRISPR-Cas9 system (IDT) and reverse transfection of a Cas9:guideRNA ribonucleoprotein complex using Lipofactamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's protocol. The crRNA sequences to target CASC3 were /AlTR1/rGrCrGrCrGrCrUrUrCrGrCrArArGrArCrArCrCrGrGrUrUrUrUrArGrArGrCrUrArUrGrCrU/AlTR2/ (clone H) and /AlTR1/rGrUrUrCrGrGrCrCrUrCrCrGrCrGrCrUrGrUrGrArGrUrUrUrUrArGrArGrCrUrArUrGrCrU/AlTR2/ (clones F and T). Reverse transfection was performed on $1.5x10^5$ cells per crRNA in 12-well dishes.

48 hours after transfection the cells were trypsinized, counted and seeded at a density of a single cell per well in 96-well plates. Cell colonies originating from a single clone were then validated by Sanger sequencing of the targeted genomic DNA locus and western blotting.

Plasmid transfection

All used plasmids are listed in Supplementary Table 1. To express FLAG-tagged protein constructs and the reporter mRNAs detected by northern blotting, the cells were stably transfected using the Flp-In T-REx system and the tetracycline inducible pcDNA5/FRT/TO vector (Thermo Fisher Scientific). The constructs TPI-WT, TPI-PTC, β-globin WT and β-globin PTC are available on Addgene (IDs 108375-108378). $2.5x10^5$ cells were seeded 24 h before transfection in 6-wells. Per well, 1 µg of reporter construct was transfected together with 1 µg of the Flp recombinase expressing plasmid pOG44 using the calcium phosphate method. 48 h after transfection, the cells were transferred into 10 cm dishes and selected with 100 µg/ml hygromycin. After 10 days, the colonies were pooled. Expression of the reporter mRNA was induced with 1 µg/ml doxycycline for 24 h.

Constructs that express V5-tagged and MS2V5-tagged proteins were stably integrated into the cells using the PiggyBac (PB) Transposon system and the cumate-inducible PB-CuO-MCS-IRES-GFP-EF1-CymR-Puro vector (System Biosciences). $2.5x10^5$ cells were seeded 24 h before transfection in 6-wells. 2.5 µg of the PB Transposon vector and 0.8 µg of PB Transposase were transfected per well using the calcium phosphate method. After 48 h, the cells were pooled in 10 cm dishes and positive clones selected with 2 µg/ml puromycin for a week. Expression of proteins was induced using 30 µg/ml cumate for 72 h.

The tethering construct pSBtet-Hyg-TPI-4MS2-SMG5-4H was stably integrated into HeLa Tet-Off cells using the Sleeping Beauty (SB) transposon system (Mates *et al.*, 2009, Kowarz *et al.*, 2015). pSBtet-Hyg was a gift from Eric Kowarz (Addgene plasmid #60508; http://n2t.net/addgene:60508; RRID:Addgene_60508). pCMV(CAT)T7-SB100 was a gift from Zsuzsanna Izsvak (Addgene plasmid #34879; http://n2t.net/addgene:34879; RRID:Addgene_34879). $2.5x10^5$ cells were seeded 24 h before

435 transfection in 6-wells. Per well, 1 µg of the reporter construct was transfected together with 1.5 µg
436 of the SB Transposase using the calcium phosphate method. 48 h after transfection, the cells were
437 transferred into 10 cm dishes and selected with 100 µg/ml hygromycin. After 10 days, the colonies
438 were pooled. In absence of tetracycline the reporter was constitutively expressed.

439 *RNA-Sequencing and computational analyses*

440 RNA-Seq analysis was carried out with 293 WT cells transfected with Luciferase siRNA and the
441 CASC3 KO clones H and T transfected with either Luciferase or CASC3 siRNAs. Three biological
442 replicates were analyzed for each sample. RNA was isolated with the kit NucleoSpin RNA Plus
443 (Macherey-Nagel). The Lexogen SIRV Set1 Spike-In Control Mix (SKU: 025.03) that provides a set of
444 external RNA controls was added to the total RNA to enable performance assessment. Mix 0 was added
445 to replicate 1, mix 1 was added to replicate 2 and mix 3 to replicate 3. The Spike-Ins were not used for
446 analysis. The library preparation was performed with the TrueSeq Stranded Total RNA kit (Illumina).
447 First steps of the library preparation involve the removal of ribosomal RNA using biotinylated target-
448 specific oligos combined with Ribo-Zero gold rRNA removal beads from 1ug total RNA input. The Ribo-
449 Zero Human/Mouse/Rat kit depletes samples of cytoplasmic and mitochondrial rRNA. Following
450 purification, the RNA is fragmented and cleaved. RNA fragments are copied into first strand cDNA using
451 reverse transcriptase and random primers, followed by second strand cDNA synthesis using DNA
452 Polymerase I and RNase H. These cDNA fragments then have the addition of a single'A' base and
453 subsequent ligation of the adapter. The products are purified and enriched with PCR to create the final
454 cDNA library. After library validation and quantification (Agilent tape station), equimolar amounts of
455 library were pooled. The pool was quantified by using the Peqlab KAPA Library Quantification Kit and
456 the Applied Biosystems 7900HT Sequence Detection System and sequenced on an Illumina
457 NovaSeq6000 sequencing instrument and an PE100 protocol.

458 Read processing and alignment was performed as described previously (Boehm *et al.*, 2018). In
459 short, adaptor sequences and low quality bases were removed with Flexbar 3.0 (Dodt *et al.*, 2012).

460 Short reads from the rRNA locus were subtracted by mapping against the 45S precursor (Homo
461 sapiens, NR_046235.1) using Bowtie2 (Langmead & Salzberg, 2012). The remaining reads were aligned
462 against the human genome (version 38, EnsEMBL 90 transcript annotations) using the STAR read
463 aligner (version 2.5.3a) (Dobin *et al.*, 2013).

464 To compute gene differential expression analysis, reads covering exons were counted with
465 FeatureCounts (version 1.5.1) (Liao *et al.*, 2014) using the '—primary' and '—ignoreDup' parameters.
466 Differential gene expression analysis was performed with DESeq2 (Love *et al.*, 2014, Ignatiadis *et al.*,
467 2016) and IWH R packages. Significance thresholds were |log2FoldChange|> 1 and adjusted p-value
468 (padj) < 0.05. Genes were designated as small RNA (sRNA) host gene, if they contained other Ensembl-
469 annotated genes of biotypes snoRNA or miRNA within their genomic coordinates (Zerbino *et al.*, 2018).

470 Differential splicing was detected with LeafCutter (version 0.2.7) (Li *et al.*, 2018) with the
471 parameters min_samples_per_intron = 2 and min_samples_per_group = 2. Significance thresholds
472 were |deltapsi| > 0.1 and adjusted p-value (p.adjust) < 0.05.

473 Transcript abundance estimates were computed with Salmon (version 0.13.1) (Patro *et al.*, 2017)
474 using the the -validateMappings --gcBias parameters. Differential transcript usage was computed with
475 IsoformSwitchAnalyzeR (version 1.7.1) and the DEXSeq method (Robinson & Oshlack, 2010, Anders *et*
476 *al.*, 2012, Ritchie *et al.*, 2015, Soneson *et al.*, 2015, Vitting-Seerup & Sandelin, 2017, Vitting-Seerup &
477 Sandelin, 2019). Significance thresholds were |dIF| > 0.1 and adjusted p-value
478 (isoform_switch_q_value) < 0.05. For the Boxplot and Kolmogorov-Smirnoff test, the data were filtered
479 only for the adjusted p-value. PTC status of transcript isoforms with annotated open reading frame
480 was determined by IsoformSwitchAnalyzeR using the 50 nt rule of NMD (Weischenfeldt *et al.*, 2012,
481 Vitting-Seerup *et al.*, 2014, Huber *et al.*, 2015, Vitting-Seerup & Sandelin, 2017). Isoforms with no
482 annotated open reading frame in Ensembl were designated "NA" in the PTC analysis.

483 The UPF1 and SMG6/7 (Data ref: Colombo *et al.*, 2017) and RNPS1 (Data ref: Boehm *et al.*, 2018)
484 knockdown datasets were processed and analyzed with the same programs, program versions, and

485 scripts as the CASC3 dataset. All packages used are listed in the respective analysis table
486 (Supplementary Tables 4-6). Sashimi plots were generated using ggsashimi (Garrido-Martin *et al.*,
487 2018). Overlaps of data sets were represented via nVenn (Perez-Silva *et al.*, 2018), eulerr (Larsson,
488 2019) and Upset plots (Lex *et al.*, 2014). Heatmaps were generated using ComplexHeatmap (Gu *et al.*,
489 2016).

490 *SILAC, co-immunoprecipitation and mass spectrometry*

491 293 WT and 293 CASC3 KO clone H cells expressing either FLAG or FLAG-EIF4A3 were labeled by
492 maintaining them for 5 passages in DMEM for SILAC medium (Thermo Fisher Scientific) supplemented
493 with FBS (Silantes), Penicillin-Streptomycin (Thermo Fisher Scientific) and the respective amino acids
494 at a final concentration of 0.798 mmol/L (Lysine) and 0.398 (Arginine). Unlabeled proline was added
495 to prevent enzymatic Arginine-to-Proline conversion. The conditions were “light” (unlabeled
496 Lysine/Arginine), “medium” (Lysine 4/Arginine 6) and “heavy” (Lysine 8/Arginine 10). A label switch
497 was performed between the three replicates according to the experimental setup listed in
498 Supplementary Table 2. The expression of FLAG or FLAG-EIF4A3 was induced for 72 h with 1 µg/ml
499 doxycycline. The cells were lysed in buffer E with RNAse (20 mM HEPES-KOH (pH 7.9), 100 mM KCl,
500 10% glycerol, 1 mM DTT, Protease Inhibitor, 1 µg/ml RNAse A) and sonicated using the Bandelin
501 Sonopuls mini20 with 15 pulses (2.5 mm tip in 600 µl volume, 1s, 50% amplitude). 600 µl of a 1.6 mg/ml
502 total protein lysate were incubated with 30 µl Anti-FLAG M2 magnetic beads (Sigma) at 4° C while
503 rotating for 2 h. The beads were washed three times for 5 min with EJC-buffer (20 mM HEPES-KOH (pH
504 7.9), 200 mM NaCl, 2 mM $MgCl_2$, 0.2% Triton X-100, 0.1% NP-40, 0.05% Sodium deoxycholate) and
505 eluted in 43 µl of a 200 mg/ml dilution of FLAG peptides (Sigma) in 1x TBS. The samples were merged
506 according to Supplementary Table 2. 1 volume of 10% SDS was added and the samples were reduced
507 with DTT and alkylated with CAA (final concentrations 5 mM and 40 mM, respectively). Tryptic protein
508 digestion was performed using a modified version of the single pot solid phase-enhanced sample
509 preparation (SP3) (Hughes *et al.*, 2014). In brief, reduced and alkylated proteins were supplemented
510 with paramagnetic Sera-Mag speed beads (Thermo Fisher Scientific) and mixed in a 1:1-ratio with 100%

511 acetonitrile (ACN). After 8 min incubation protein-beads-complexes were captured using an in-house
512 build magnetic rack and two times washed with 70% EtOH. Afterwards, samples were washed once
513 with 100% ACN, air-dried and reconstituted in 5 µl 50 mM Triethylamonium bicarbonate
514 supplemented with 0.5 µg trypsin and 0.5 µg LysC and incubated overnight at 37°C. On the next day
515 the beads were resuspended and mixed with 200 µl ACN, incubated for 8 min and again placed on the
516 magnetic rack. Tryptic peptides were washed once with 100% ACN, airdried, dissolved in 4% DMSO
517 and transferred into 96-well PCR tubes. After acidification with 1 µl of 10% formic acid, samples were
518 ready for LC-MS/MS analysis

519 For the MS experiment combined with CASC3 knockdown, the 293 CASC3 KO clone H cells were
520 treated with siRNA against CASC3 24 h before expression of the FLAG-tagged construct. The cells were
521 lysed and samples prepared for MS as described above. However, the washing steps were performed
522 in a less stringent EJC buffer (20 mM HEPES-KOH (pH 7.9), 137 mM NaCl, 2 mM $MgCl_2$, 0.2% Triton X-
523 100, 0.1% NP-40).

524 Proteomics analysis was performed by data-dependent acquisition using an Easy nLC1200 ultra
525 high-performance liquid chromatography (UHPLC) system coupled via nanoelectrospray ionization to
526 a Q Exactive Plus instrument (all Thermo scientific). Tryptic peptides were separated based on their
527 hydrophobicity using a chromatographic gradient of 60 min with a binary system of buffer A (0.1%
528 formic acid) and buffer B (80% ACN, 0.1% formic acid). In-house made analytical columns (length: 50
529 cm, inner diameter: 75 µm) filled with 1.9 µm C18-AQ Reprosil Pur beads (Dr. Maisch) were used for
530 separation. Buffer B was linearly increased from 3% to 27% over 41 min followed by a steeper increase
531 to 50% within 8 min. Finally, buffer B was increased to 95% within 1 min and stayed at 95% for 10 min
532 in order to wash the analytical column. Full MS spectra (300 – 1,750 m/z) were acquired with a
533 resolution of 70,000, a maximum injection time of 20 ms and an AGC target of 3e6. The top 10 most
534 abundant peptide ions of each full MS spectrum were selected for HCD fragmentation (NCE: 27) with

an isolation width of 1.8 m/z and a dynamic exclusion of 10 seconds. MS/MS spectra were measured with a resolution of 35,000, a maximum injection time of 110 ms and an AGC target of 5e5.

MS RAW files were analysed using the standard settings of the MaxQuant suite (version 1.5.3.8) with the before mentioned SILAC labels (Cox & Mann, 2008). Peptides were identified by matching against the human UniProt database using the Andromeda scoring algorithm (Cox *et al.*, 2011). Carbamidomethylation of cysteine was set as a fixed modification, methionine oxidation and N-terminal acetylation as variable modification. Trypsin/P was selected as digestion protein. A false discovery Rate (FDR) < 0.01 was used for identification of peptide-spectrum matches and protein quantification. Data processing and statistical analysis was done in the Perseus software (version 1.5.5.3) (Tyanova *et al.*, 2016). Significantly changed proteins were identified by One-sample t-testing (H0 = 0, fudge factor S0 = 0.1). The results are listed in Supplementary Table 2 as "Results without knockdown" (MAP conditions) and "Results with knockdown". Visualization was performed with the Instant Clue software (version 0.5.3) (Nolte *et al.*, 2018).

Co-immunoprecipitation experiments followed by western blotting were performed as described above except that a 15 min incubation step in SDS buffer (600 mM Tris pH 6.8, 100 mM DTT, 10% Glycerol, 2% SDS, 0.002% Bromophenolblue) was used for elution from the beads.

In vitro splicing assay

FLAG-emGFP and FLAG-EIF4A3 was expressed in 293 WT and 293 CASC3 KO clone H cells for 72 h. Whole cell lysate preparation and *in vitro* splicing was performed as previously described (Steckelberg & Gehring, 2014), however the nuclear extract was substituted with whole cell extract from the respective cell line. Pulldown efficiency of spliced mRNA was calculated by dividing the amount of spliced mRNA in the pulldown by the amount in the respective input and dividing the results by 10 since 10% of the input was loaded.

Semi-quantitative and quantitative reverse transcriptase (RT)-PCR

559 RNA was extracted using peqGOLD TriFast reagent (VWR) according to the manufacturer's
560 instructions. Reverse transcription was performed with GoScript Reverse Transcriptase (Promega)
561 using 2 µg total RNA and oligo dT primers. Semi-quantitative PCR was carried out with MyTaq Red Mix
562 (Bioline). Quantitative real time PCR was performed with 16 ng of cDNA per reaction with GoTaq qPCR
563 Master Mix (Promega) and the CFX96 Touch Real-Time PCR Detection System (Biorad). The average cT
564 values were calculated from three technical replicates. The mean fold changes from three biological
565 replicates were calculated according to the ΔΔCt method (Schmittgen & Livak, 2008). When measuring
566 isoform switches, the fold change of the PTC-containing transcript was normalized to the canonical
567 transcript. When measuring differential expression, the fold change was normalized to GAPDH. For
568 each primer pair amplification efficiencies were measured by a 2-fold dilution curve and ranged
569 between 87 and 100.1%. The primer sequences are listed in Supplementary Table 3.

570 *Western blotting*

571 Protein extraction was performed with peqGOLD TriFast reagent (VWR), separated by SDS-PAGE
572 gel electrophoresis and transferred to a PVDF membrane (GE Healthcare Life Sciences). The following
573 antibodies were used: Anti-β-actin (Sigma-Aldrich, #A5441), anti-CASC3 amino acid residues 653-703
574 (Bethyl Laboratories, #A302-472A-M), anti-CASC3 amino acid residues 367-470 (Atlas Antibodies,
575 #HPA024592), anti-EIF4A3 (Genscript), anti-FLAG (Cell Signaling Technology, #14793), anti-MAGOH
576 (Santa Cruz Biotechnology, #sc-271365), anti-RBM8A (Atlas Antibodies, #HPA018403), anti-SMG6
577 (Abcam, #ab87539), anti-SMG7 (Elabscience, #E-AB-32926), anti-Tubulin (Sigma-Aldrich, #T6074), anti-
578 V5 (QED Bioscience, #18870), anti-XRN1 (Bethyl Laboratories, #A300-443A), anti-rabbit-HRP (Jackson
579 ImmunoResearch, #111-035-006), anti-mouse-HRP (Jackson ImmunoResearch, #115-035-003).
580 Detection was performed with Western Lightning Plus-ECL (PerkinElmer) or Amersham ECL prime (Ge
581 Healthcare Life Sciences) and the chemiluminescence imager Fusion FX6 EDGE (Vilber-Lourmat).

582 *Northern blotting*

The cells were harvested in peqGOLD TriFast reagent (VWR) and total RNA extraction was performed as recommended by the manufacturer's protocol. 2.5 µg of total RNA were resolved on a 1% agarose/0.4 M formaldehyde gel using the tricine/triethanolamine buffer system (Mansour & Pestov, 2013) followed by transfer on a nylon membrane (Roth) in 10x SSC. The blots were incubated overnight at 65 °C in Church buffer containing [α-32P]-GTP body-labeled RNA probes for detection of the reporter mRNA. Endogenous 7SL RNA was detected by a 5'-32P-labeled oligonucleotide (5'-TGCTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTT-3'). The blots were visualized and quantified using a Typhoon FLA 7000 (GE Healthcare) and ImageQuant TL 1D software.

Data Availability

The datasets produced in this study are available in the following databases. These data will be made publicly accessible upon publication.

- RNA-seq data have been deposited in the ArrayExpress database (Kolesnikov *et al.*, 2015) at EMBL-EBI under accession number E-MTAB-8461 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8461).
- The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al.*, 2019) partner repository with the dataset identifier PXD015754 (https://www.ebi.ac.uk/pride/archive/projects/PXD015754).

Author Contributions

Conceptualization, N.H.G., J.V.G. and V.B.;

Methodology, N.H.G., V.B., J.V.G., and C.K.F.;

Software, T.B.B., V.B., J.L.W., S.K. and C.D.;

Investigation, J.V.G, V.B., J.L.W., S.K., D.U.A. and S.C.;

Resources and Data Curation, T.B.B., J.L.W., S.K., J.A. and C.D.;

Writing – Original Draft, Review & Editing, N.H.G., J.V.G. and V.B.;

Visualization, J.V.G., V.B. and T.B.B.;

Supervision, N.H.G., C.D. and M.K.;

Funding Acquisition, N.H.G. and C.D.

Acknowledgements

We thank members of the Gehring lab for discussions and reading of the manuscript. We also thank Marek Franitza and Christian Becker (Cologne Center for Genomics, CCG) for preparing the sequencing libraries and operating the sequencer. We acknowledge Tobias Jakobi for helping with infrastructure support. This work was supported by grants from the Deutsche Forschungsgemeinschaft to C.D. (DI 1501/8-1, DI1501/8-2) and N.H.G (GE 2014/6-1, GE 2014/6-2 and GE 2014/10-1). V.B. was funded under the Institutional Strategy of the University of Cologne within the German Excellence Initiative. N.H.G. acknowledges support by a Heisenberg fellowship (GE 2014/5-1 and GE 2014/7-1) from the Deutsche Forschungsgemeinschaft. C.D. and T.B.B. were kindly supported by the Klaus Tschira Stiftung gGmbH (00.219.2013)

Conflict of Interest

The authors declare no competing interests.

References

Alexandrov A, Colognori D, Shu MD, Steitz JA (2012) Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc Natl Acad Sci U S A* 109: 21313-8

Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* 22: 2008-17

Andersen CB, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CL, Pedersen JS, Seraphin B, Le Hir H, Andersen GR (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313: 1968-72

Ballut L, Marchadier B, Baguet A, Tomasetto C, Seraphin B, Le Hir H (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol* 12: 861-9

Barbosa I, Haque N, Fiorini F, Barrandon C, Tomasetto C, Blanchette M, Le Hir H (2012) Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat Struct Mol Biol* 19: 983-90

Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R (2011) The quantitative proteome of a human cell line. *Mol Syst Biol* 7: 549

Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452: 846-50

Boehm V, Britto-Borges T, Steckelberg AL, Singh KK, Gerbracht JV, Gueney E, Blazquez L, Altmuller J, Dieterich C, Gehring NH (2018) ArrayExpress E-MTAB-6564 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6564)

Boehm V, Britto-Borges T, Steckelberg AL, Singh KK, Gerbracht JV, Gueney E, Blazquez L, Altmuller J, Dieterich C, Gehring NH (2018) Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. *Mol Cell* 72: 482-495 e7

Boehm V, Gehring NH (2016) Exon Junction Complexes: Supervising the Gene Expression Assembly Line. *Trends Genet* 32: 724-735

Boehm V, Gerbracht JV, Marx MC, Gehring NH (2016) Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences. *Nat Commun* 7: 13691

Boehm V, Haberman N, Ottens F, Ule J, Gehring NH (2014) 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep* 9: 555-68

Bono F, Ebert J, Lorentzen E, Conti E (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126: 713-25

Buchwald G, Ebert J, Basquin C, Sauliere J, Jayachandran U, Bono F, Le Hir H, Conti E (2010) Insights into the recruitment of the NMD machinery from the crystal structure of a core EJC-UPF3b complex. *Proc Natl Acad Sci U S A* 107: 10050-5

Buhler M, Steiner S, Mohn F, Paillusson A, Muhlemann O (2006) EJC-independent degradation of nonsense immunoglobulin-mu mRNA depends on 3' UTR length. *Nat Struct Mol Biol* 13: 462-4

661 Colombo M, Karousis ED, Bourquin J, Bruggmann R, Muhlemann O (2017) Gene Expression Omnibus
662 GSE86148 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE86148). [DATASET]

663 Colombo M, Karousis ED, Bourquin J, Bruggmann R, Muhlemann O (2017) Transcriptome-wide
664 identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and
665 SMG7-mediated degradation pathways. *RNA* 23: 189-201

666 Cougot N, Daguenet E, Baguet A, Cavalier A, Thomas D, Bellaud P, Fautrel A, Godey F, Bertrand E,
667 Tomasetto C, Gillet R (2014) Overexpression of MLN51 triggers P-body disassembly and formation of
668 a new type of RNA granules. *J Cell Sci* 127: 4692-701

669 Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range
670 mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26: 1367-72

671 Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M (2011) Andromeda: a peptide search
672 engine integrated into the MaxQuant environment. *J Proteome Res* 10: 1794-805

673 Degot S, Regnier CH, Wendling C, Chenard MP, Rio MC, Tomasetto C (2002) Metastatic Lymph Node
674 51, a novel nucleo-cytoplasmic protein overexpressed in breast cancer. *Oncogene* 21: 4422-34

675 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013)
676 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21

677 Dodt M, Roehr JT, Ahmed R, Dieterich C (2012) FLEXBAR-Flexible Barcode and Adapter Processing for
678 Next-Generation Sequencing Platforms. *Biology (Basel)* 1: 895-905

679 Durand S, Lykke-Andersen J (2013) Nonsense-mediated mRNA decay occurs during eIF4F-dependent
680 translation in human cells. *Nat Struct Mol Biol* 20: 702-9

681 Eberle AB, Lykke-Andersen S, Muhlemann O, Jensen TH (2009) SMG6 promotes endonucleolytic
682 cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* 16: 49-55

683 Garrido-Martin D, Palumbo E, Guigo R, Breschi A (2018) ggsashimi: Sashimi plot revised for browser-
684 and annotation-independent splicing visualization. *PLoS Comput Biol* 14: e1006360

685 Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH, Hentze MW, Kulozik AE (2005) Exon-junction
686 complex components specify distinct routes of nonsense-mediated mRNA decay with differential
687 cofactor requirements. *Mol Cell* 20: 65-75

688 Gehring NH, Lamprinaki S, Hentze MW, Kulozik AE (2009) The hierarchy of exon-junction complex
689 assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay.
690 *PLoS Biol* 7: e1000120

691 Gerbracht JV, Gehring NH (2018) The exon junction complex: structural insights into a faithful
692 companion of mammalian mRNPs. *Biochem Soc Trans* 46: 153-161

693 Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in
694 multidimensional genomic data. *Bioinformatics* 32: 2847-9

695 Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M,
696 Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland OS, Durocher D, Angers S, Moffat
697 J (2015) High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities.
698 *Cell* 163: 1515-26

699 Hart T, Tong AHY, Chan K, Van Leeuwen J, Seetharaman A, Aregger M, Chandrashekhar M, Hustedt N,
700 Seth S, Noonan A, Habsid A, Sizova O, Nedyalkova L, Climie R, Tworzyanski L, Lawson K, Sartori MA,
701 Alibeh S, Tieu D, Masud S et al. (2017) Evaluation and Design of Genome-Wide CRISPR/SpCas9
702 Knockout Screens. *G3 (Bethesda)* 7: 2719-2727

703 Hentze MW, Castello A, Schwarzl T, Preiss T (2018) A brave new world of RNA-binding proteins. *Nat*
704 *Rev Mol Cell Biol* 19: 327-341

705 Hoek TA, Khuperkar D, Lindeboom RGH, Sonneveld S, Verhagen BMP, Boersma S, Vermeulen M,
706 Tanenbaum ME (2019) Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA
707 Decay. *Mol Cell* 75: 324-339 e11

708 Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke
709 T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oles
710 AK, Pages H et al. (2015) Orchestrating high-throughput genomic analysis with Bioconductor. *Nat*
711 *Methods* 12: 115-21

712 Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J (2014) Ultrasensitive proteome
713 analysis using paramagnetic bead technology. *Mol Syst Biol* 10: 757

714 Huntzinger E, Kashima I, Fauser M, Sauliere J, Izaurralde E (2008) SMG6 is the catalytic endonuclease
715 that cleaves mRNAs containing nonsense codons in metazoan. *RNA* 14: 2609-17

716 Hurt JA, Robertson AD, Burge CB (2013) Global analyses of UPF1 binding and function reveal expanded
717 scope of nonsense-mediated mRNA decay. *Genome Res* 23: 1636-50

718 Ignatiadis N, Klaus B, Zaugg JB, Huber W (2016) Data-driven hypothesis weighting increases detection
719 power in genome-scale multiple testing. *Nat Methods* 13: 577-80

720 Ishigaki Y, Li X, Serin G, Maquat LE (2001) Evidence for a pioneer round of mRNA translation: mRNAs
721 subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* 106:
722 607-17

723 Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S (2006)
724 Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1
725 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev* 20: 355-67

726 Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, Wang F, Dreyfuss G (2000) Pre-mRNA splicing
727 imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell*
728 6: 673-82

729 Kolesnikov N, Hastings E, Keays M, Melnichuk O, Tang YA, Williams E, Dylag M, Kurbatova N, Brandizi
730 M, Burdett T, Megy K, Pilicheva E, Rustici G, Tikhonov A, Parkinson H, Petryszak R, Sarkans U, Brazma
731 A (2015) ArrayExpress update--simplifying data submissions. *Nucleic Acids Res* 43: D1113-6

732 Kowarz E, Loscher D, Marschalek R (2015) Optimized Sleeping Beauty transposons rapidly generate
733 stable transgenic cell lines. *Biotechnol J* 10: 647-53

734 Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357-9

735 Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE (2007) Unproductive splicing of SR genes
736 associated with highly conserved and ultraconserved DNA elements. *Nature* 446: 926-9

737 Larsson J (2019) eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version
738 6.0.0, https://cran.r-project.org/package=eulerr

739 Le Hir H, Gatfield D, Braun IC, Forler D, Izaurralde E (2001) The protein Mago provides a link between
740 splicing and mRNA localization. *EMBO Rep* 2: 1119-24

741 Le Hir H, Sauliere J, Wang Z (2016) The exon junction complex as a node of post-transcriptional
742 networks. *Nat Rev Mol Cell Biol* 17: 41-54

743 Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H (2014) UpSet: Visualization of Intersecting Sets.
744 *IEEE Trans Vis Comput Graph* 20: 1983-92

745 Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK (2018) Annotation-free
746 quantification of RNA splicing using LeafCutter. *Nat Genet* 50: 151-158

747 Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning
748 sequence reads to genomic features. *Bioinformatics* 30: 923-30

749 Loh B, Jonas S, Izaurralde E (2013) The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT
750 deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* 27:
751 2125-38

752 Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq
753 data with DESeq2. *Genome Biol* 15: 550

754 Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, Jensen TH (2014) Human nonsense-
755 mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev*
756 28: 2498-517

757 Mabin JW, Woodward LA, Patton RD, Yi Z, Jia M, Wysocki VH, Bundschuh R, Singh G (2018) The Exon
758 Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-
759 Mediated mRNA Decay Activity. *Cell Rep* 25: 2431-2446 e7

760 Mansour FH, Pestov DG (2013) Separation of long RNA by agarose-formaldehyde gel electrophoresis.
761 *Analytical biochemistry* 441: 18-20

762 Mao H, Brown HE, Silver DL (2017) Mouse models of Casc3 reveal developmental functions distinct
763 from other components of the exon junction complex. *RNA* 23: 23-31

764 Maquat LE, Tarn WY, Isken O (2010) The pioneer round of translation: features and functions. *Cell* 142:
765 368-74

766 Mates L, Chuah MK, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, Grzela DP, Schmitt A, Becker K,
767 Matrai J, Ma L, Samara-Kuko E, Gysemans C, Pryputniewicz D, Miskey C, Fletcher B, VandenDriessche
768 T, Ivics Z, Izsvak Z (2009) Molecular evolution of a novel hyperactive Sleeping Beauty transposase
769 enables robust stable gene transfer in vertebrates. *Nat Genet* 41: 753-61

770 Metze S, Herzog VA, Ruepp MD, Muhlemann O (2013) Comparison of EJC-enhanced and EJC-
771 independent NMD in human cells reveals two partially redundant degradation pathways. *RNA* 19:
772 1432-48

773 Nagy E, Maquat LE (1998) A rule for termination-codon position within intron-containing genes: when
774 nonsense affects RNA abundance. *Trends Biochem Sci* 23: 198-9

775 Nasif S, Contu L, Muhlemann O (2018) Beyond quality control: The role of nonsense-mediated mRNA
776 decay (NMD) in regulating gene expression. *Semin Cell Dev Biol* 75: 78-87

777 Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M, Jr.
778 (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by
779 alternative splicing and nonsense-mediated decay. *Genes Dev* 21: 708-18

780 Nolte H, MacVicar TD, Tellkamp F, Kruger M (2018) Instant Clue: A Software Suite for Interactive Data
781 Visualization and Analysis. *Sci Rep* 8: 12648

782 Palacios IM, Gatfield D, St Johnston D, Izaurralde E (2004) An eIF4AIII-containing complex required for
783 mRNA localization and nonsense-mediated mRNA decay. *Nature* 427: 753-7

784 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C (2017) Salmon provides fast and bias-aware
785 quantification of transcript expression. *Nat Methods* 14: 417-419

786 Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J,
787 Mayer G, Eisenacher M, Perez E, Uszkoreit J, Pfeuffer J, Sachsenberg T, Yilmaz S, Tiwary S, Cox J, Audain
788 E, Walzer M, Jarnuczak AF et al. (2019) The PRIDE database and related tools and resources in 2019:
789 improving support for quantification data. *Nucleic Acids Res* 47: D442-D450

790 Perez-Silva JG, Araujo-Voces M, Quesada V (2018) nVenn: generalized, quasi-proportional Venn and
791 Euler diagrams. *Bioinformatics* 34: 2322-2324

792 Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential
793 expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47

794 Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of
795 RNA-seq data. *Genome Biol* 11: R25

796 Rufener SC, Muhlemann O (2013) eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA
797 decay in mammalian cells. *Nat Struct Mol Biol* 20: 710-7

798 Schmidt SA, Foley PL, Jeong DH, Rymarquis LA, Doyle F, Tenenbaum SA, Belasco JG, Green PJ (2015)
799 Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of
800 endogenous NMD targets in human cells. *Nucleic Acids Res* 43: 309-23

801 Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nature*
802 *protocols* 3: 1101-8

803 Shibuya T, Tange TO, Sonenberg N, Moore MJ (2004) eIF4AIII binds spliced mRNA in the exon junction
804 complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* 11: 346-51

805 Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ (2012) The cellular EJC
806 interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* 151: 750-764

807 Soneson C, Love MI, Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates
808 improve gene-level inferences. *F1000Res* 4: 1521

809 Steckelberg AL, Altmueller J, Dieterich C, Gehring NH (2015) CWC22-dependent pre-mRNA splicing and
810 eIF4A3 binding enables global deposition of exon junction complexes. *Nucleic Acids Res* 43: 4687-700

811 Steckelberg AL, Boehm V, Gromadzka AM, Gehring NH (2012) CWC22 connects pre-mRNA splicing and
812 exon junction complex assembly. *Cell Rep* 2: 454-61

Steckelberg AL, Gehring NH (2014) Studying the composition of mRNPs in vitro using splicing-competent cell extracts. *Methods* 65: 342-9

Sureau A, Gattoni R, Dooghe Y, Stevenin J, Soret J (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J* 20: 1785-96

Tani H, Imamachi N, Salam KA, Mizutani R, Ijiri K, Irie T, Yada T, Suzuki Y, Akimitsu N (2012) Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. *RNA Biol* 9: 1370-9

Thermann R, Neu-Yilik G, Deters A, Frede U, Wehr K, Hagemeier C, Hentze MW, Kulozik AE (1998) Binary specification of nonsense codons by splicing and cytoplasmic translation. *EMBO J* 17: 3484-94

Trcek T, Sato H, Singer RH, Maquat LE (2013) Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev* 27: 541-51

Tuladhar R, Yeu Y, Tyler Piazza J, Tan Z, Rene Clemenceau J, Wu X, Barrett Q, Herbert J, Mathews DH, Kim J, Hyun Hwang T, Lum L (2019) CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat Commun* 10: 4056

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* 13: 731-40

Vitting-Seerup K, Porse BT, Sandelin A, Waage J (2014) spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* 15: 81

Vitting-Seerup K, Sandelin A (2017) The Landscape of Isoform Switches in Human Cancers. *Mol Cancer Res* 15: 1206-1220

Vitting-Seerup K, Sandelin A (2019) IsoformSwitchAnalyzeR: Analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*

Voigt F, Gerbracht JV, Boehm V, Horvathova I, Eglinger J, Chao JA, Gehring NH (2019) Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts. *Nature protocols* 14: 1603-1633

Wang Z, Murigneux V, Le Hir H (2014) Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol* 15: 551

Weischenfeldt J, Waage J, Tian G, Zhao J, Damgaard I, Jakobsen JS, Kristiansen K, Krogh A, Wang J, Porse BT (2012) Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol* 13: R35

Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol Cell* 13: 91-100

Woodward LA, Mabin JW, Gangras P, Singh G (2017) The exon junction complex: a lifelong guardian of mRNA fate. *Wiley Interdiscip Rev RNA* 8

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Giron CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR et al. (2018) Ensembl 2018. *Nucleic Acids Res* 46: D754-D761

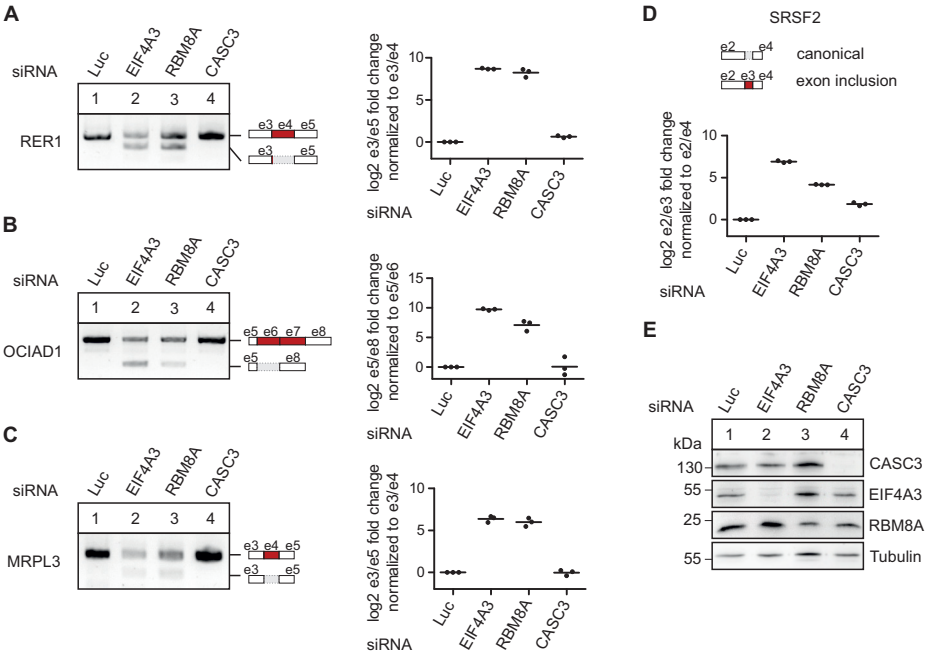


Figure 1 - CASC3 is not involved in the splicing regulation of known EJC-dependent targets.

A-C: RT-PCR- and quantitative RT-PCR-based detection (qPCR) of transcript isoforms of the genes RER1 (A), OCIAD1 (B), and MRPL3 (C) after siRNA-mediated knockdown of the indicated EJC components or Luciferase (Luc) as a negative control. Skipped exons are depicted schematically (e: exon). Data points and means from the qPCRs are plotted (n=3).

D: Relative quantification of the SRSF2 transcript isoforms by qPCR following knockdown of the indicated EJC components or Luciferase (Luc) as a negative control. The transcript variants at the position of the included exon are depicted schematically. Data points and means are plotted (n=3).

E: Confirmation of the knockdowns shown in A-D by western blotting.

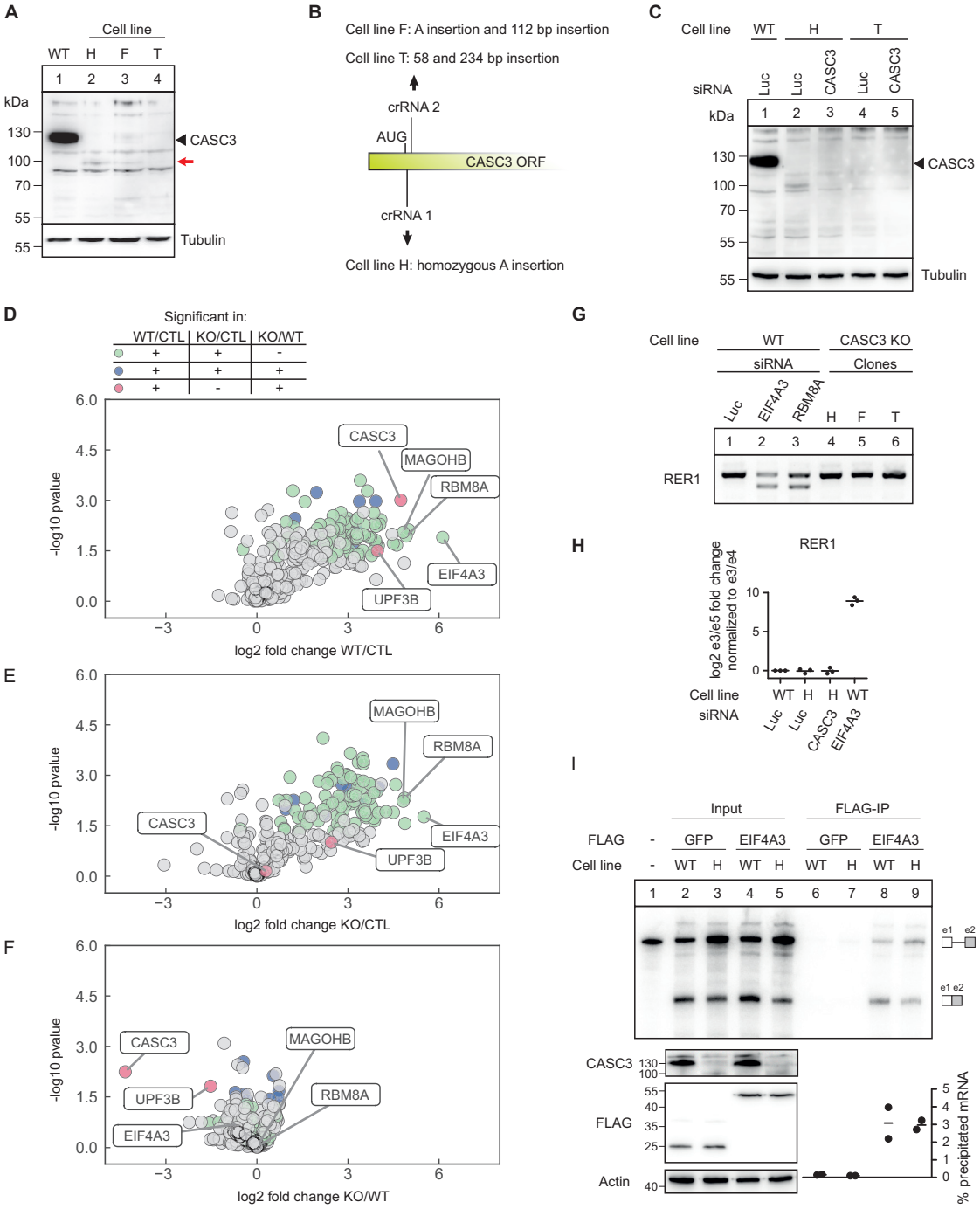


Figure 2 - Cells that lack CASC3 have intact EJCs and are splicing competent.

A: Total protein lysates from wild-type cells (WT) and CASC3 knockout (KO) cell lines H, F and T were separated by SDS-PAGE and CASC3 was detected by western blotting. The red arrow indicates an additional band detected in the cell lines H and F and which is not visible in the WT condition or in the cell line T.

B: Schematic depiction of the insertions resulting in a CASC3 KO or constitutive knockdown in the indicated clones.

C: Total protein lysates from WT and CASC3 KO cell lines H and T were separated by SDS-PAGE and CASC3 was detected by western blotting. In lanes 3 and 5 the cells have additionally been treated with siRNAs targeting CASC3.

D-F: Volcano plots of mass spectrometry-based analysis of the interaction partners of EIF4A3 in WT cells and in the CASC3 KO cell line H treated with siRNAs targeting CASC3. D: EIF4A3 against FLAG control in WT cells, E: EIF4A3 against FLAG control in KO cells, F: EIF4A3 in KO cells against EIF4A3 in WT cells. The color labeling indicates targets that are significant in the respective comparisons after one-sample t-testing.

G: RT-PCR of transcript isoforms of the gene RER1 after siRNA-mediated knockdown of the indicated EJC components or Luciferase (Luc) as a negative controle, compared to CASC3 KO cell lines H, F and T.

H: Relative quantification of the RER1 transcript isoforms by qPCR in WT cells treated with Luc siRNA as a negative control, CASC3 KO cell line H treated with Luc siRNA, CASC3 KO cell line H treated with CASC3 siRNAs and WT cells treated with EIF4A3 siRNA. Data points and means are plotted (n=3).

I: *In vitro* splicing assay of a MINX transcript in cell lysates obtained from WT cells or the CASC3 KO cell line H which stably expressed FLAG-GFP or FLAG-EIF4A3. The protein levels of CASC3 and the recombinant proteins in the cell lysates were analyzed by western blotting. The pulldown efficiencies of spliced mRNA were calculated and are plotted as individual data points and means (n=2).

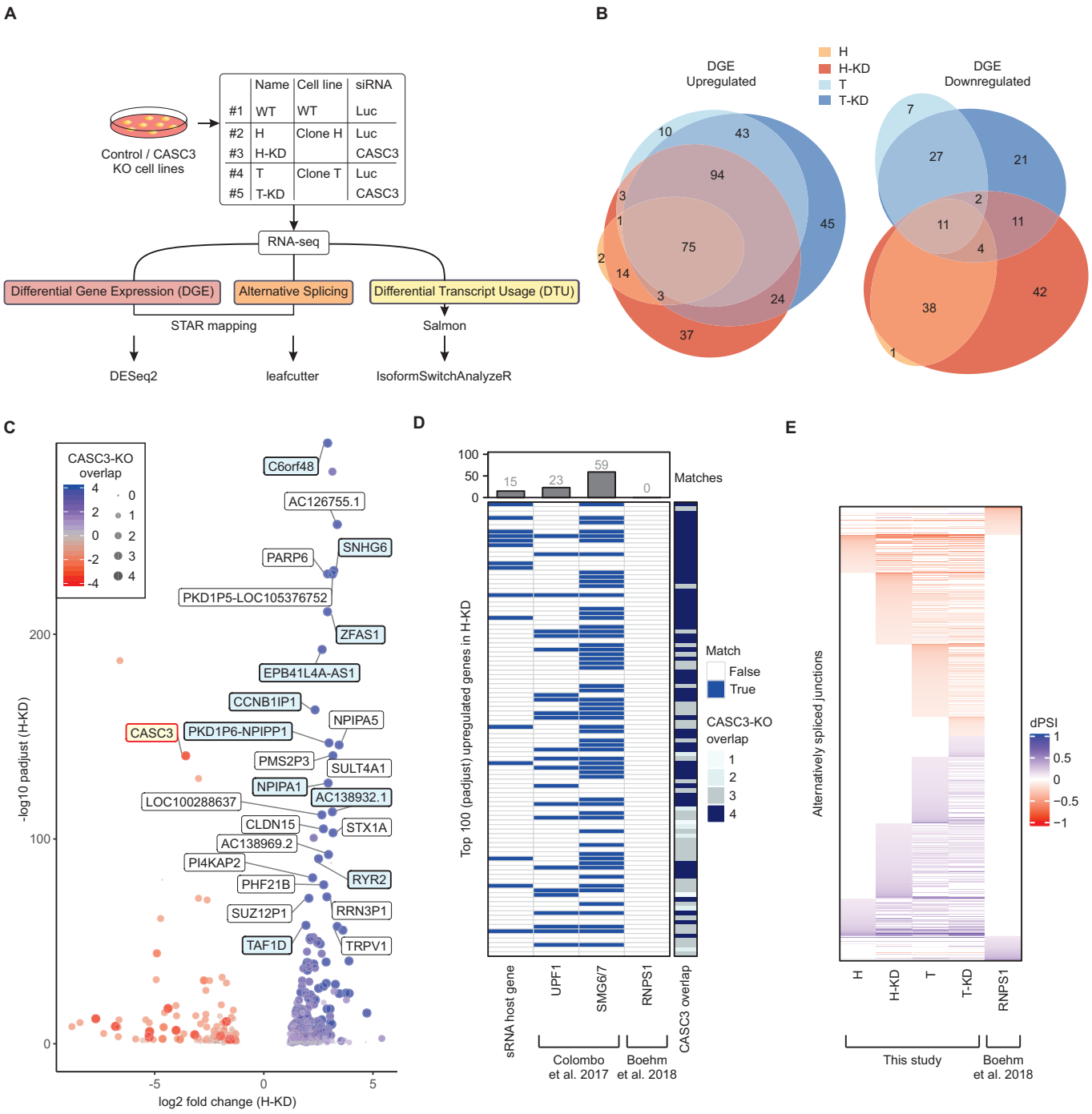


Figure 3 - CASC3-dependent targets are also affected by a perturbation of the NMD machinery.

A: Pipeline for RNA-sequencing analysis.

B: Overlap of up- and downregulated genes in the CASC3 KO cell lines H and T, +/- CASC3 siRNAs. DGE: Differential gene expression. Due to the visualization as Euler plot, some intersections cannot be plotted. All intersections are shown in Supplementary Figure 3C.

C: Volcano plot of differential gene expression analysis of the condition H-KD using overlap from B as color and point size definition. Gene symbols are indicated for the top 25 upregulated genes detected in all four conditions and for CASC3 (colored in light red). Labels of small RNA host genes are colored in light blue. Log2 fold change is plotted against –log10 padjust (adjusted p-value).

D: Matching of the top 100 upregulated genes sorted by padjust (adjusted p-value) in condition H-KD with small RNA (sRNA) host genes and comparison to knockdowns of UPF1, SMG6/7 and RNPS1.

E: Heatmap of all identified alternatively spliced junctions in the respective condition, measured in delta percent spliced in (dPSI).

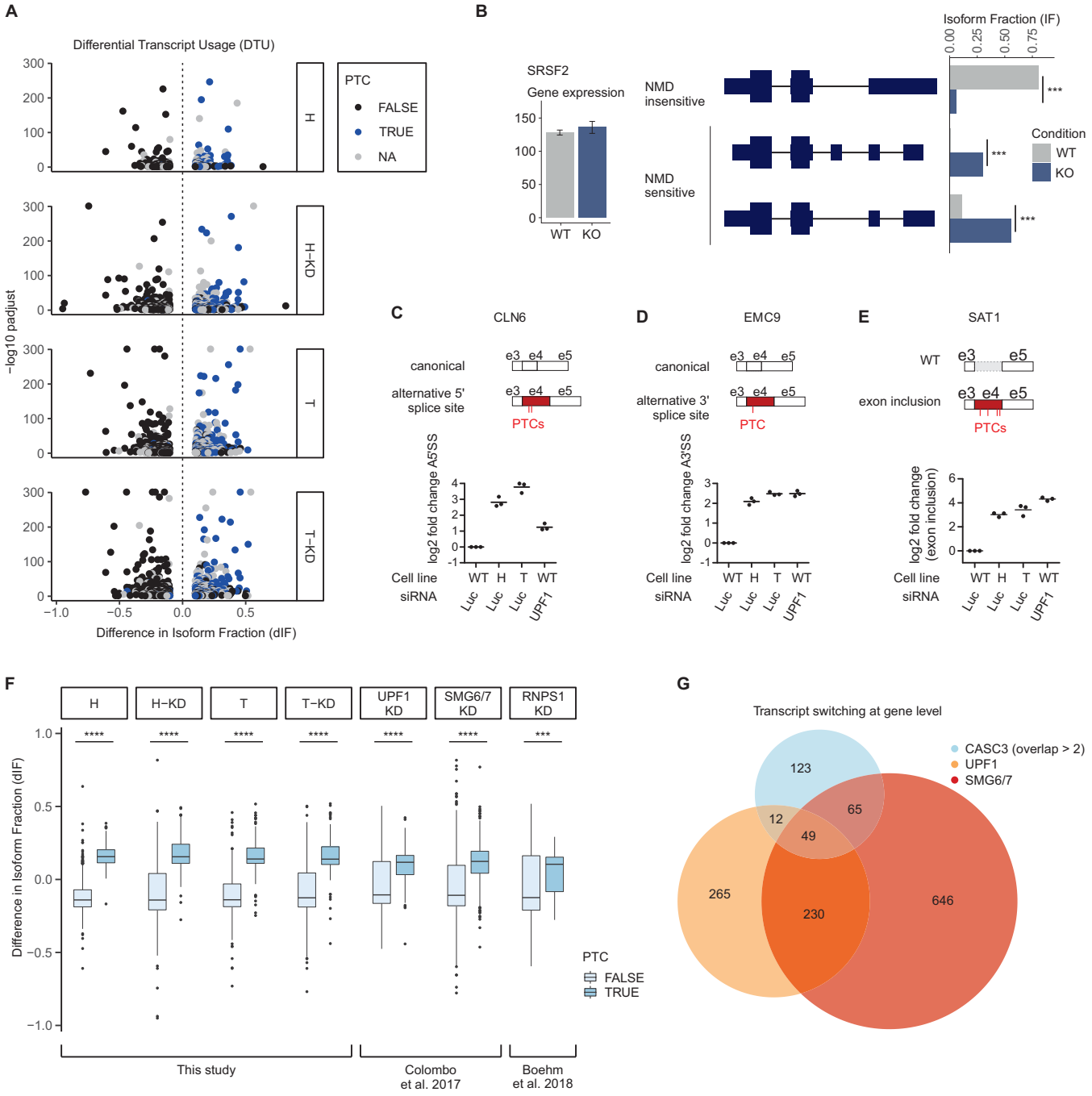


Figure 4 - Knockout of CASC3 leads to a global upregulation of NMD-sensitive transcript isoforms.

A: Results from IsoformSwitch analysis plotted as volcano plots. Transcript isoforms identified as NMD-sensitive are shown as blue dots. Isoforms with no annotated open reading frame are designated as "NA". Difference in Isoform Fraction (dIF) is plotted against –log10 padjust (adjusted p-value).

B: Quantification of transcript isoforms from SRSF2 by IsoformSwitchAnalyzeR.

C-E: Relative quantification of the schematically depicted transcript isoforms of the genes CLN6 (C), EMC9 (D), and SAT1 (E) by qPCR in WT cells, CASC3 KO cell lines H and T and WT cells treated with siRNA targeting UPF1. PTC: premature termination codon. Individual data points and means are plotted (n=3).

F: Boxplot of PTC-containing vs. non-PTC-containing transcript isoforms after IsoformSwitch analysis for all CASC3 KO conditions compared to UPF1, SMG6/7 and RNPS1. A Kolmogorov-Smirnoff test was applied (p-value < 0.001 ***, p-value < 10-16 ****).

G: Euler plot showing overlaps of DTU-affected genes in the CASC3 (> 2 overlaps), UPF1-KD and SMG6/7-KD condition.

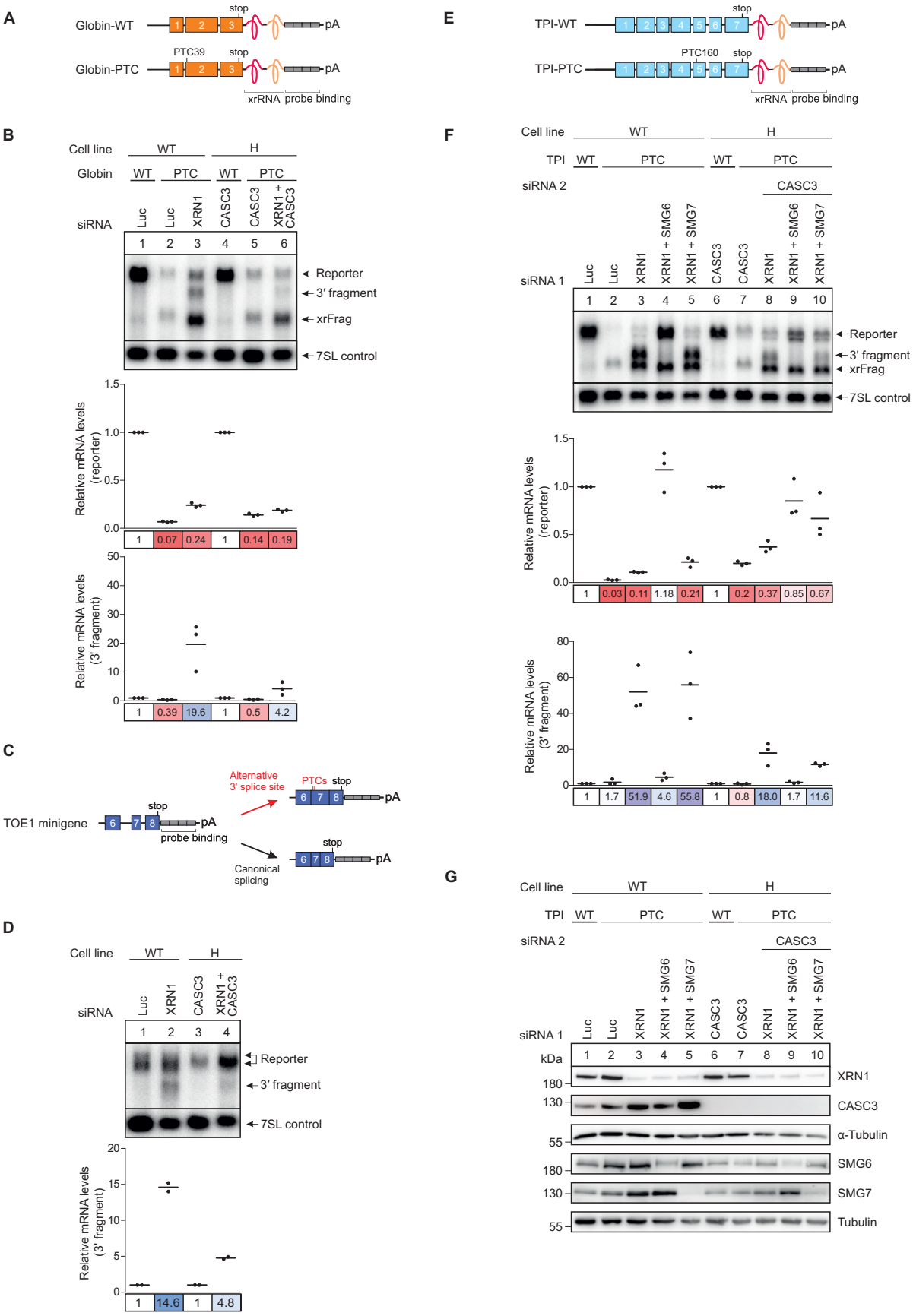


Figure 5 - SMG6-mediated endocleavage is impaired when CASC3 is not present.

A: Schematic depiction of the globin mRNA reporter. The reporter consists of three exons (orange boxes) followed by an XRN1-resistant element (xrRNA) and a probe binding cassette (gray boxes). The PTC reporter contains a premature termination codon (PTC) in the second exon.

B: Northern blot of RNA extracted from the indicated cell lines that stably express the globin reporter. The xrFrag corresponds to the 3′ part of the reporter that is resistant to degradation by XRN1 due to the xrRNA. The cell lines in lane 3 and 6 were additionally treated with XRN1 siRNA which results in the appearance of a 3′ degradation fragment below the full-length reporter. Reporter and 3′ fragment mRNA levels were normalized to 7SL RNA which is shown as a loading control. For the relative mRNA quantification, in each condition (WT vs. CASC3 KO with KD) the reporter and 3′ fragment levels were normalized to the globin WT reporter (lanes 1 and 4). Individual data points and means are plotted from n=3 experiments.

C: Schematic depiction of the TOE1 minigene reporter consisting of exons 6-8 (purple boxes) followed by a probe binding cassette (gray boxes). The reporter can be spliced to either contain the canonical stop codon (bottom right) or, by usage of an alternative 3′ splice site, a PTC in exon 7 (top right).

D: Northern blot of RNA extracted from the indicated cell lines treated with the indicated siRNAs stably expressing the TOE1 minigene reporter. The 3′ fragment levels were first normalized to the 7SL RNA loading control and for every cell line the XRN1 knockdown condition to the condition without XRN1 knockdown (n=2).

E: Schematic depiction of the triose phosphate isomerase (TPI) mRNA reporter. The reporter consists of seven exons (blue boxes) followed by an XRN1-resistant element (xrRNA) and a probe binding cassette (gray boxes). The PTC reporter contains a premature termination codon (PTC) in the fifth exon.

F: Northern blot of RNA extracted from the indicated cell lines treated with the indicated siRNAs stably expressing the either the TPI WT or TPI PTC mRNA reporter. The reporter and 3′ fragment mRNA levels were normalized to the 7SL control. For each cell line, the mRNA levels were then normalized to the respective TPI WT reporter or 3′ fragment levels. Individual data points and means are plotted from n=3 experiments.

G: Western blot of samples shown in Figure 5F.

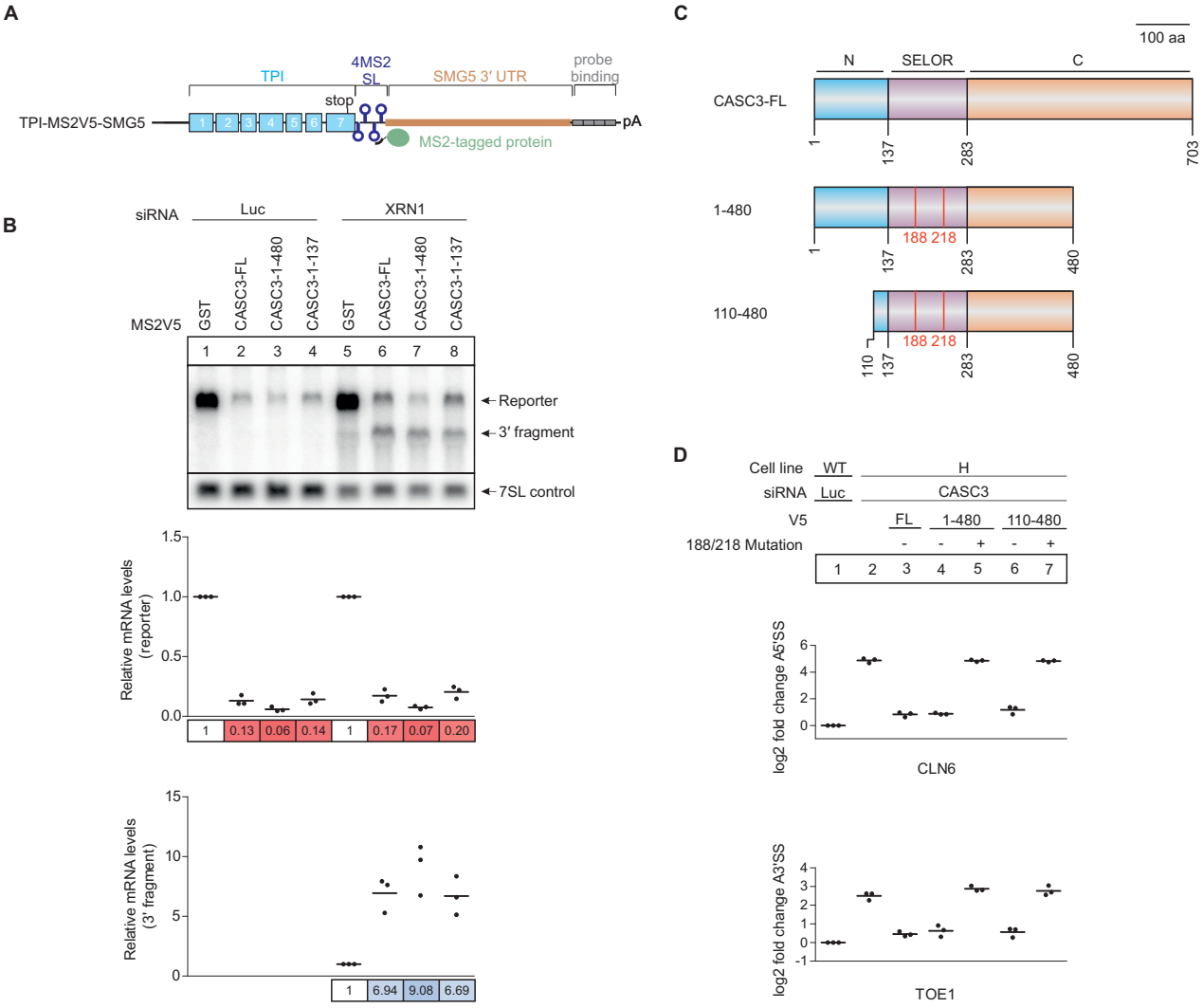


Figure 6 - The CASC3 N-terminus promotes but is not necessary to elicit NMD.

A: Schematic depiction of the TPI-MS2V5-SMG5 tethering reporter. The reporter consists of the TPI ORF (blue boxes) followed by 4 MS2 stem loops (SL). Downstream the SMG5 3′ untranslated region (UTR) is inserted to increase the size of 3′ fragments that result from cleavage at the termination codon. Reporter and 3′ fragment mRNAs can be detected via the probe binding cassette (gray boxes).

B: Northern blot of a tethering assay performed in HeLa Tet-Off cells. The cells stably express the tethering reporter shown in A together with the indicated MS2V5-tagged proteins. When the cells are additionally treated with XRN1 siRNA, a 3′ degradation fragment can be detected below the full-length reporter. The reporter and 3′ fragment mRNA levels are normalized to the 7SL RNA. For the calculation of the relative mRNA levels in each condition (Luc vs. XRN1) the levels were normalized to the MS2V5-GST control (lanes 1 and 5).

C: Schematic depiction of CASC3 rescue protein constructs. The full-length (FL) protein consists of an N-terminal (blue), C-terminal (orange) and central SELOR domain (purple). The construct 1-480 has a C-terminal deletion, whereas in the construct 110-480 both the N- and C-terminus are truncated. Both deletion constructs were also rendered EJC-binding deficient by mutating the amino acid residues 188 and 218 (F188D, W218D).

D: Relative quantification of the CLN6 (top) and TOE1 (bottom) transcript isoforms by qPCR in the indicated cell lines. The V5-tagged rescue proteins expressed in the KO condition are shown in C. Individual data points and means are plotted from n=3 experiments.

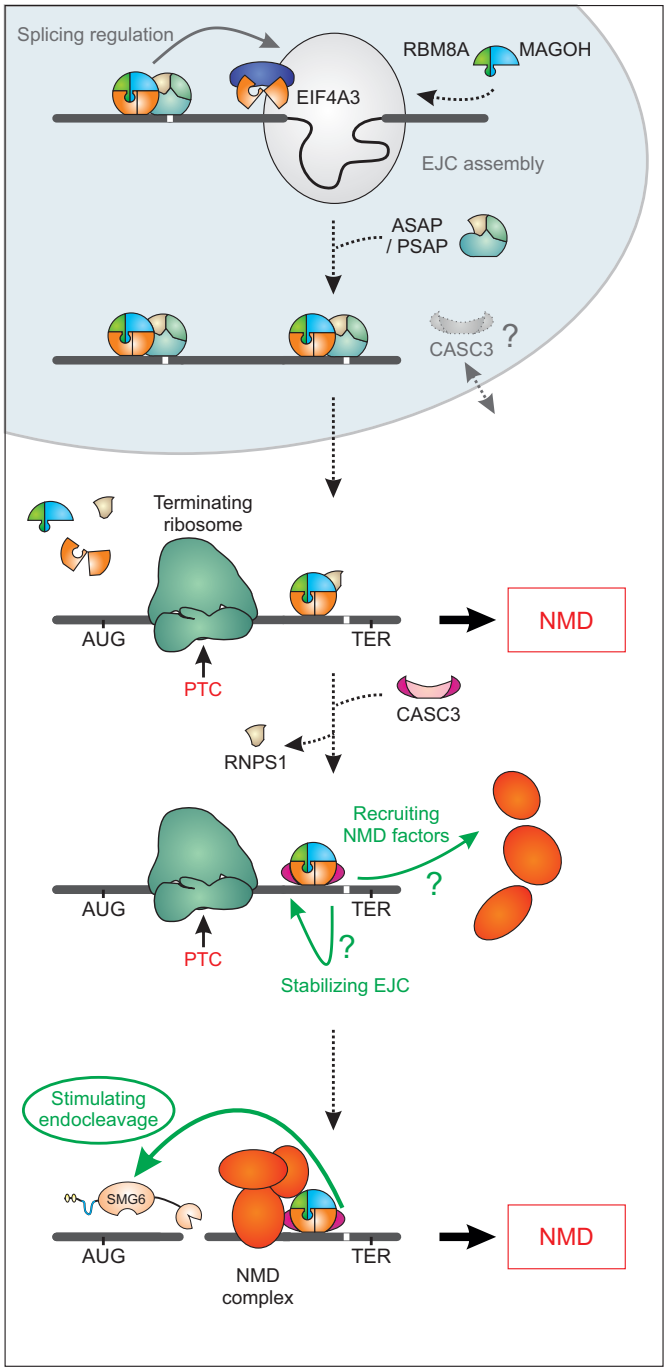
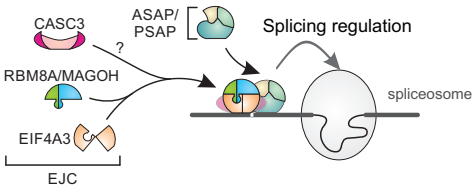


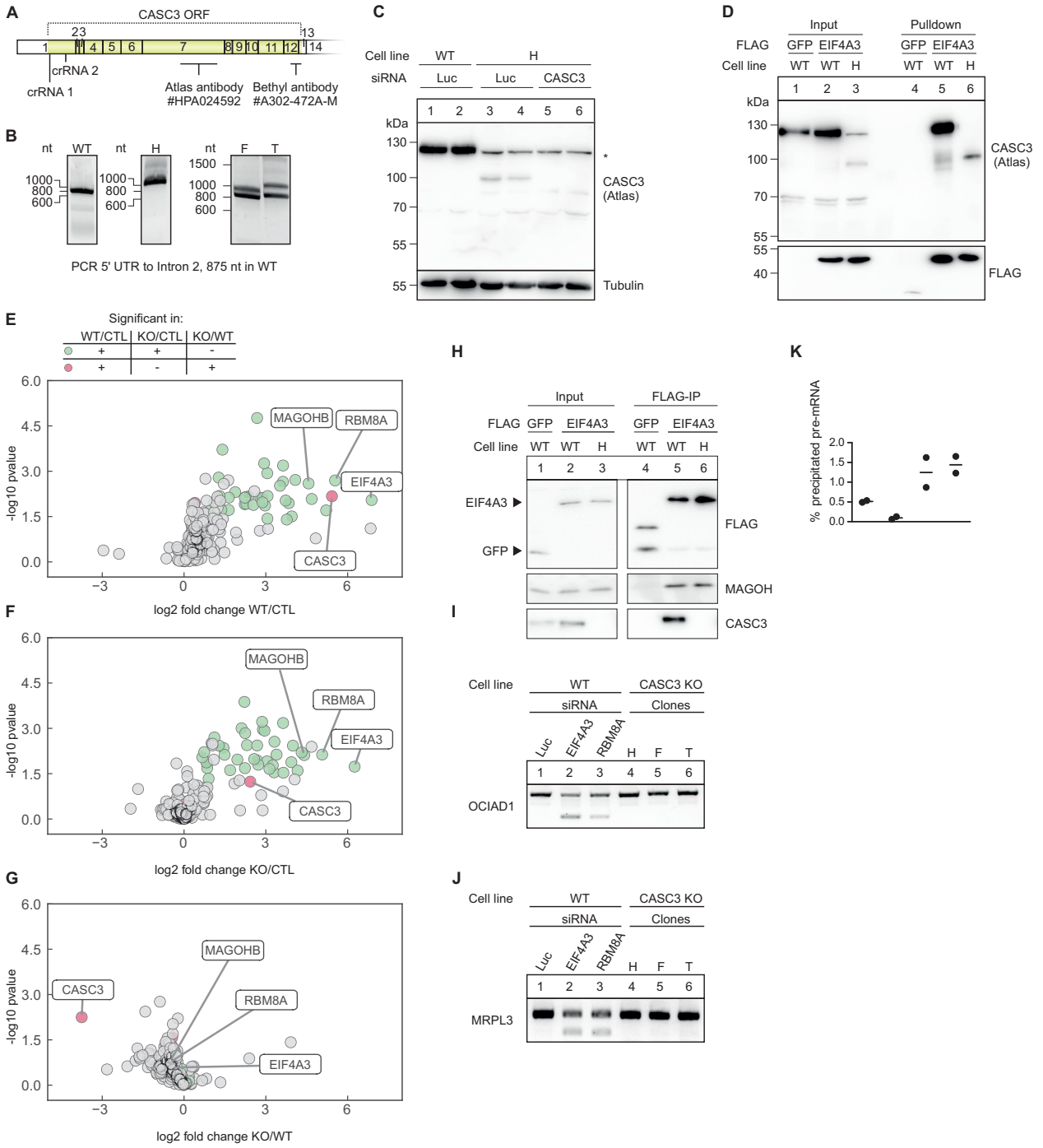
Figure 7 - Model of CASC3's cellular function.

Schematic depiction of the proposed function of CASC3 in the cell. We have found no evidence that CASC3 is necessary for EJC assembly in the nucleus, although CASC3 shuttles between cytoplasm and nucleus. Transcripts that require the deposition of the EJC to be correctly spliced were not affected by a lack of CASC3. In the cytoplasm, premature termination codon (PTC)-containing transcripts may still be degraded by NMD during the initial round(s) of translation in a CASC3-independent manner. CASC3 association with the EJC maintains and/or promotes the NMD-stimulating effect of the EJC, resulting in the degradation of transcripts that evaded initial NMD activation.



Supplementary Figure 1 - Model of splicing regulation by the EJC.

Modified from Boehm *et al.* (2018).



Supplementary Figure 2 - EJC composition in CASC3 KO cells.

A: Recognition sites of the two antibodies used in this study to detect CASC3. The coding exons 1-13 and a part of the non-coding exon 14 of CASC3 are depicted. The open reading frame (ORF) is marked yellow. The positions of the crRNA sequences to generate the knockout cells are also indicated.

B: Amplification of the genomic CASC3 locus targeted by CRISPR-Cas9 editing in the cell lines H, F and T. The sequences of the insertions were analyzed by cloning and Sanger-sequencing of the respective PCR-fragments and are shown in Figure 2B.

C: Western blot of cell lysates from WT cells and the CASC3 KO cell line H that have been treated with the indicated siRNAs.

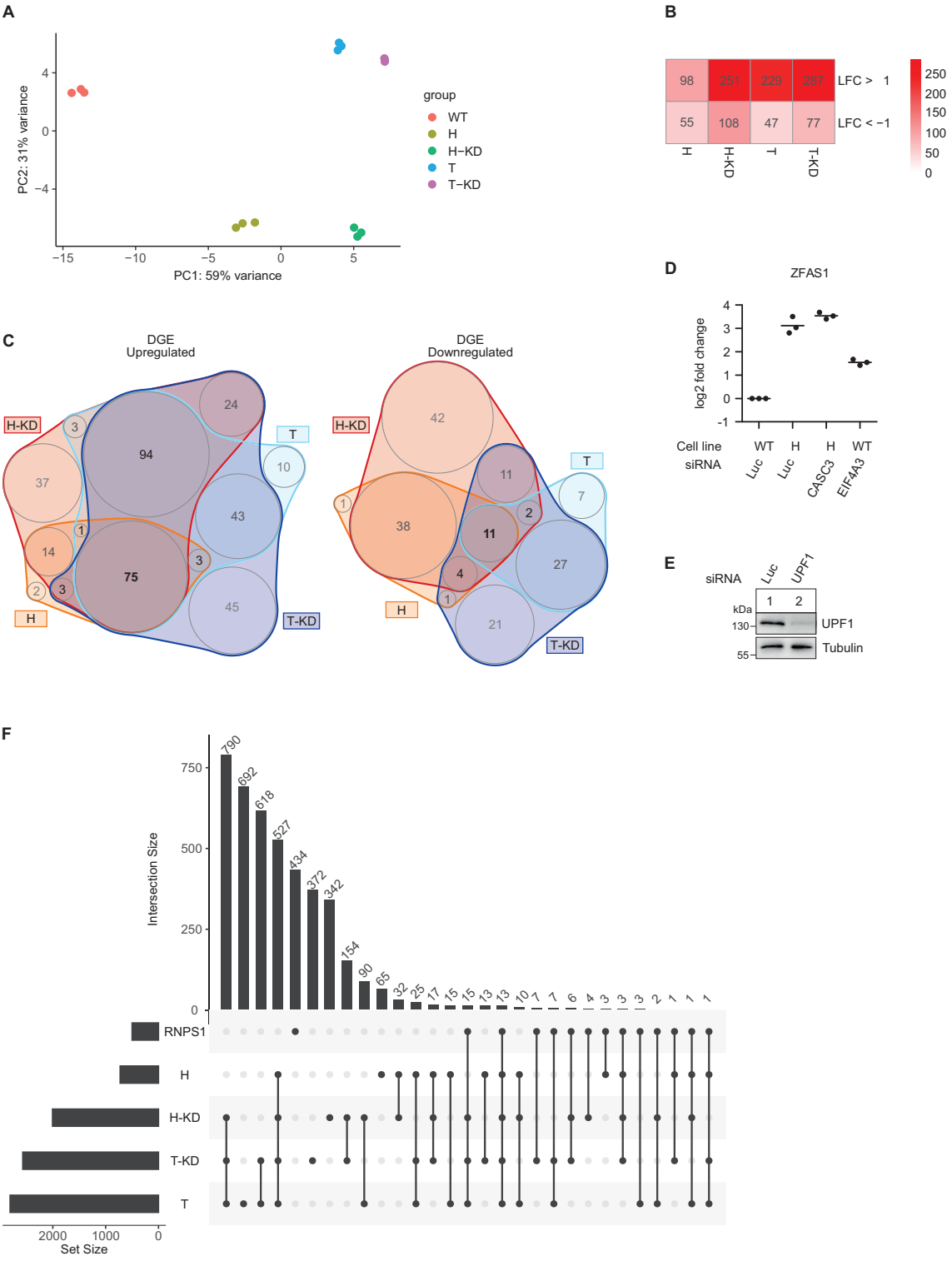
D: Co-immunoprecipitation of FLAG-GFP and FLAG-EIF4A3 that were stably expressed in the indicated cell lines. Endogenous CASC3 and the recombinant FLAG-tagged proteins were detected by western blotting.

E-G: Volcano plots of mass spectrometry-based analysis of the interaction partners of EIF4A3 in WT cells and in the CASC3 KO cell line H. E: EIF4A3 against FLAG control in WT cells, F: EIF4A3 against FLAG control in KO cells, G: EIF4A3 in KO cells against EIF4A3 in WT cells. The color labeling indicates targets that are significant in the respective comparisons after one-sample t-testing.

H: Co-immunoprecipitation of FLAG-GFP and FLAG-EIF4A3 that were stably expressed in the indicated cell lines. Endogenous CASC3 and MAGOH as well as the recombinant FLAG proteins are detected by western blotting.

I and J: RT-PCR of transcript isoforms of the genes OCIAD1 (I) and MRPL3 (J) after siRNA-mediated knockdown of the indicated EJC components, Luciferase (Luc) as a negative control and in the CASC3 KO cell lines H, F and T.

K: Relative amounts of precipitated pre-mRNA in the *in vitro* splicing assay shown in Figure 2I.



Supplementary Figure 3 - Transcriptome-wide effects of the CASC3 KO.

A: Principal component analysis (PCA) of differential gene expression data.

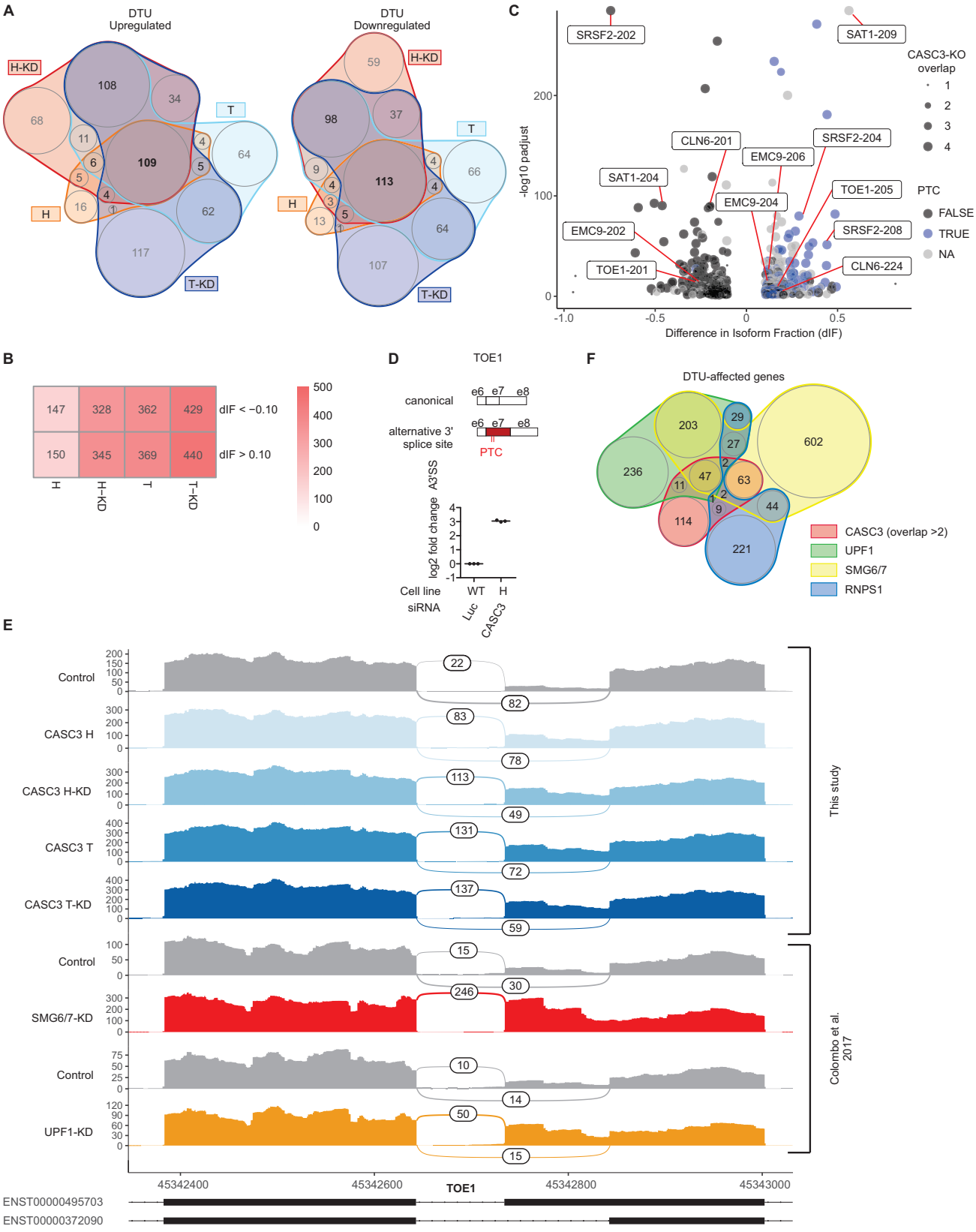
B: Summary of significantly up- or downregulated genes in the differential gene expression analysis with an absolute log2 fold change (|LFC|) > 1.

C: Overlap of all intersections of differentially expressed genes in all CASC3 KO conditions shown as nVenn plot.

D: Relative quantification by qPCR of the ZFAS1 transcript in WT cells, CASC3 KO cell lines H and T and WT cells treated with siRNA targeting UPF1. Individual data points and means are plotted (n=3).

E: Confirmation of the UPF1 knockdown by western blotting.

F: Upset plot of intersections of alternatively spliced junctions in the shown conditions.



Supplementary Figure 4 - Analysis of differentially expressed transcript isoforms in the CASC3 KO.

A: Overlap of all intersections of transcript isoforms up-or downregulated in IsoformSwitch analysis shown as nVenn plot. DTU: Differential transcript usage.

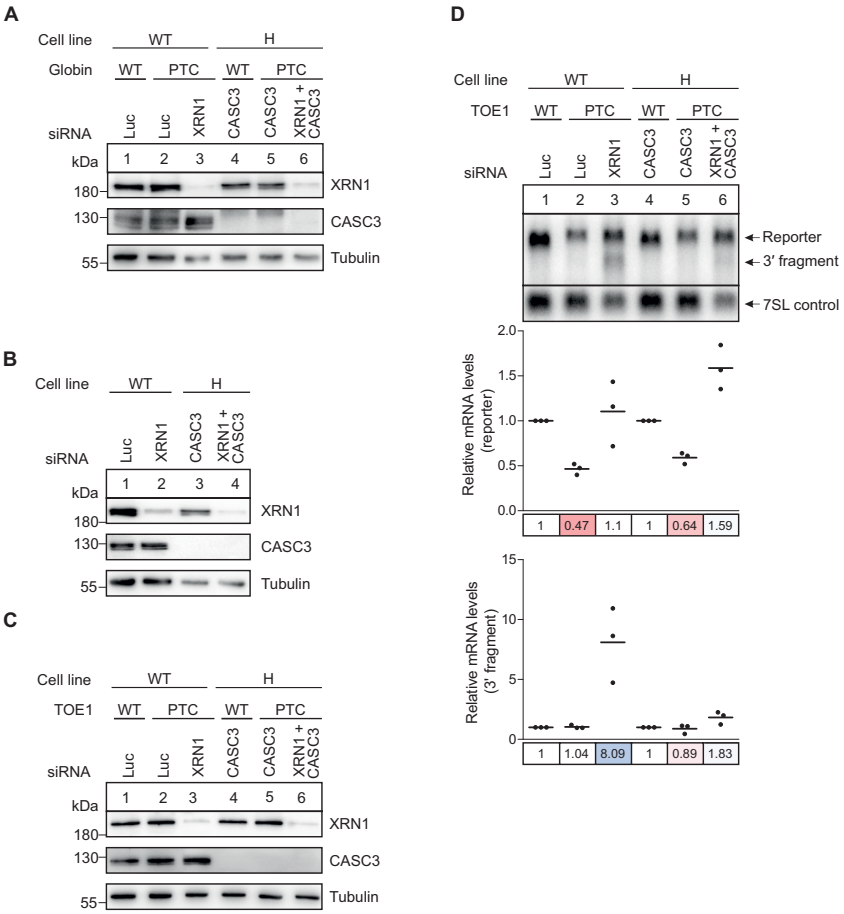
B: Summary of transcript isoforms up- or downregulated in IsoformSwitch analysis with an absolute Difference in Isoform Fraction (|dIF|) > 0.1.

C: Volcano plot of IsoformSwitch data for the condition H-KD using overlap from Figure S4A as point size definition and PTC status as color definition. Isoforms with no annotated open reading frame are designated as "NA". Ensembl transcript names are indicated for the genes validated in Figures 4C-E and S4D. Difference in Isoform Fraction (dIF) is plotted against –log10 padjust (adjusted p-value).

D: Relative quantification of the schematically depicted transcript isoforms from the gene TOE1 by qPCR in WT cells and CASC3 KO cell line H treated with CASC3 siRNAs. PTC: premature termination codon. Individual data points and means are plotted (n=3).

E: Sashimi plots of the CASC3 KO RNA-seq dataset compared to data from Colombo *et al.* (2017) for the gene TOE1. The counts of reads spanning the indicated junctions are shown.

F: Overlap of DTU-affected genes in conditions CASC3 (> 2 overlaps), UPF1-KD, SMG6/7-KD and RNPS1-KD shown as nVenn plot.



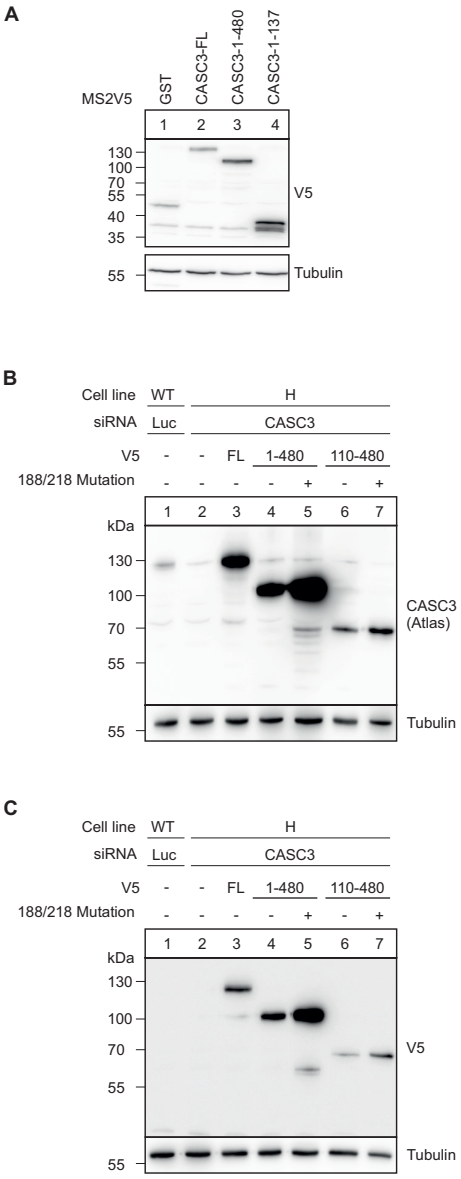
Supplementary Figure 5 - Degradation of a prespliced TOE1 minigene reporter in WT and CASC3 KO cells.

A: Western blot of samples shown in Figure 5B.

B: Western blot of samples shown in Figure 5D.

C: Western blot of samples shown in Supplementary Figure 5D.

D: Northern blot of RNA extracted from the indicated cell lines treated with the indicated siRNAs stably expressing a prespliced variant of the TOE1 minigene reporter depicted in Figure 5C. The intron between exon 6 and 7 is deleted so that the reporter is constitutively spliced to contain either a normal stop codon (TOE1-WT) or a PTC (TOE1-PTC). The mRNA levels were normalized to 7SL RNA. For the relative mRNA quantification, in each condition (WT vs. CASC3 KO with KD) the reporter and 3′ fragment levels were normalized to the TOE-WT reporter (lanes 1 and 4).



Supplementary Figure 6 - Confirmation of tethering- and rescue protein expression.

A: Western blot of samples shown in Figure 6B.

B and C: Western blot of samples shown in Figure 6D. The expression of rescue proteins was confirmed by an antibody against CASC3 (B) and an antibody recognizing the V5 tag (C).

nature COMMUNICATIONS

ARTICLE

Received 23 Feb 2016 | Accepted 25 Oct 2016 | Published 5 Dec 2016

DOI: 10.1038/ncomms13691 OPEN

Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences

Volker Boehm[1,*], Jennifer V. Gerbracht[1,*], Marie-Charlotte Marx[1] & Niels H. Gehring[1]

The turnover of messenger RNAs (mRNAs) is a key regulatory step of gene expression in eukaryotic cells. Due to the complexity of the mammalian degradation machinery, the contribution of decay factors to the directionality of mRNA decay is poorly understood. Here we characterize a molecular tool to interrogate mRNA turnover via the detection of XRN1-resistant decay fragments (xrFrag). Using nonsense-mediated mRNA decay (NMD) as a model pathway, we establish xrFrag analysis as a robust indicator of accelerated 5′–3′ mRNA decay. In tethering assays, monitoring xrFrag accumulation allows to distinguish decapping and endocleavage activities from deadenylation. Moreover, xrFrag analysis of mRNA degradation induced by miRNAs, AU-rich elements (AREs) as well as the 3′ UTRs of cytokine mRNAs reveals the contribution of 5′–3′ decay and endonucleolytic cleavage. Our work uncovers formerly unrecognized modes of mRNA turnover and establishes xrFrag as a powerful tool for RNA decay analyses.

[1] Institute for Genetics, Department of Biology, University of Cologne, Zuelpicher Straße 47a, 50674 Cologne, Germany. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to N.H.G. (email: ngehring@uni-koeln.de).

Degradation of messenger RNAs (mRNAs) directly influences the number of transcripts available for translation and thereby controls the persistence of the genetic information[1]. The regulation of mRNA degradation is widely used as a molecular principle underlying the correct expression of genes, for example of the immune system[2,3]. Many cytokine mRNAs exhibit short half-life times due to destabilizing sequences, which allow to fine-tune inflammatory responses[4,5]. In addition, the RNA decay machinery monitors the quality of mRNAs and eliminates faulty transcripts[2]. A well-studied example is the degradation of transcripts containing premature termination codons (PTCs) by the nonsense-mediated mRNA decay pathway (NMD)[6,7].

The process of mRNA decay is regulated by a large number of trans-acting protein factors. Many of them are mRNA-binding proteins, which directly interact with their target mRNAs via specific binding sites[8,9]. Destabilizing and stabilizing factors antagonistically regulate the turnover of bound mRNAs. Hence, the combination of RNA-binding proteins on a given mRNA will eventually determine the half-life of this mRNA and thereby its fate and potential for protein production.

Three major pathways of mRNA decay exist in mammalian cells: 5′–3′ exonucleolytic-, 3′–5′ exonucleolytic- and endonucleolytic decay. Deadenylation is considered to be the first and rate-limiting step during the turnover of normal cellular mRNAs[10]. Accelerated deadenylation is often initiated by the specific recruitment of the CCR4-NOT deadenylase complex, for example, by the SMG5–SMG7 heterodimer during NMD[11]. After deadenylation, further degradation of the mRNA occurs via the cytoplasmic Lsm1-7-Pat1 complex in combination with the eIF4E-binding protein 4E-T, which recruit decapping factors to the 5′ end of the deadenylated mRNA[12,13]. After decapping, the 5′–3′ exoribonuclease XRN1 recognizes the 5′ monophosphate and degrades the entire transcript[14,15]. Alternatively, the deadenylated mRNA may also be eliminated by the cytoplasmic exosome, a multi-protein complex, via its 3′–5′ exonucleolytic subunit DIS3L (ref. 16). Furthermore, transcripts with shortened poly(A) tails can be 3′-oligouridylated by TUTases, activating different degradation pathways including the 3′–5′ decay via the exosome-independent DIS3L2 (refs 17–19).

The decay of mRNA may also be initiated by endonucleolytic cleavage. A few endonucleases have been described in mammalian cells, amongst them ZC3H12A (Regnase-1) and the NMD-specific endonuclease SMG6, both of which harbour a PIN-like RNase domain[20]. Regnase-1 specifically recognizes and cleaves a stem–loop structure present in the 3′ UTRs of many cytokine mRNAs, such as TNF-α and Interleukin 6 (ref. 4). In contrast, SMG6 is recruited to PTC-containing mRNAs and cleaves them in the vicinity of the termination codon[21].

Despite the importance of mRNA turnover for the regulation of gene expression, little is known about the contribution of degradation pathways to the decay of individual mRNAs. This lack of knowledge is partially due to the large variety of seemingly redundant nucleolytic enzymes. In this study, we aimed to understand the degradation of different classes of intrinsically unstable mRNAs in mammalian cells using a virus-derived RNA sequence. Insertion of XRN1-resistant sequences (xrRNAs) into different reporter mRNAs allowed to monitor RNA decay activity of NMD substrates, cytokine 3′ UTR-containing transcripts and mRNAs containing AU-rich- or microRNAs (miRNA)-responsive elements. Using this method, we detect for these unstable transcripts a differential contribution of mRNA degradation pathways, including endocleavage, deadenylation and decapping.

Results

Monitoring NMD activity using XRN1-resistant RNA elements. Decapping or endocleavage of mRNAs generates decay intermediates with 5′ unprotected ends, which are substrates of the cytoplasmic 5′–3′ exonuclease XRN1 (Fig. 1a). The depletion of XRN1 is commonly used to detect decay intermediates and to draw conclusions about the mRNA decay mechanism. However, XRN1 interacts directly with components of the decapping machinery, suggesting that XRN1 itself modulates the decapping process[22]. Hence, we sought to develop a system to monitor 5′–3′ mRNA decay with minimal cellular invasiveness. A potential alternative to the knockdown of XRN1 is the incorporation of viral xrRNA elements in reporter mRNAs (Fig. 1a), which were previously reported to block the processively degrading XRN1 upstream of the xrRNA structure[23]. The resulting XRN1-resistant decay fragments (henceforth called xrFrag) accumulate and can be used as a readout for mRNA degradation pathways involving 5′–3′ decay (Fig. 1a).

To test the feasibility of this approach, we inserted an xrRNA element from the Murray Valley encephalitis (MVE) Virus downstream of the triosephosphate isomerase (TPI) open reading frame (ORF; Fig. 1b,c). The MVE xrRNA is structurally and molecularly well described and consists of two xrRNA sequences (xrRNA1 and xrRNA2)[23]. By northern blot analysis of transiently transfected cells we observed in addition to the control mRNA two reporter-derived RNA species. The slower migrating band represents the full-length TPI-xrRNA transcript, whereas the weak, faster migrating band corresponds in size to the expected xrFrag and likely results from regular mRNA turnover (Fig. 1d). Introducing an NMD-activating PTC (PTC160) in the TPI ORF resulted in decreased steady state reporter levels as well as increased xrFrag abundance, in line with the enhanced turnover of this mRNA (Fig. 1d, lane 2). Interestingly, no xrFrags were detected from reporter mRNAs with single xrRNA sequences (xrRNA1 or xrRNA2), establishing that two xrRNA structures (constituting one complete xrRNA element) are required for efficient XRN1 resistance. To investigate the influence of upstream RNA elements on xrRNA functionality, we inserted 60 bp sequences with varying GC contents (30–70% GC) derived from the RAB7A 3′ UTR. To the best of our knowledge, this RNA does not contain decay-inducing features and is recommended as housekeeping gene for gene expression studies[24] (Fig. 1e). To simulate more challenging potential roadblocks for XRN1, we inserted 4MS2 binding sites or a very stable stem–loop structure upstream of the xrRNA (Fig. 1e). All inserts were fully compatible with xrFrag generation (Fig. 1f), thus establishing that the xrRNA element can be readily integrated into different reporter systems.

We next asked if the xrFrag could be used to study the degradation of NMD substrates in more detail. To this end we expressed wild type (WT) or PTC-containing versions of β-globin or TPI reporter mRNAs with or without xrRNA sequences (Fig. 2a and Supplementary Fig. 1a). The presence of the xrRNA element did not alter the overall degradation efficiency of the reporter (Fig. 2b and Supplementary Fig. 1b). Furthermore, an increase of xrFrag levels was consistently detected for both PTC-containing reporters. The currently assumed major degradation pathway of NMD substrates is the SMG6-catalysed endonucleolytic cleavage at the stop codon. Of the resulting unstable 5′- and 3′ fragments, the 3′ fragments are rapidly removed by XRN1 (refs 25,26). 3′ fragments were stabilized by the knockdown of XRN1 (Fig. 2c) and appeared in addition to the xrFrag (Fig. 2b and Supplementary Fig. 1b). Depletion of the key NMD factor UPF1 abolished the difference in xrFrag:reporter ratios for the WT and PTC reporters, indicating that NMD activity is required for the increased xrFrag levels. To gain more insight into the generation of the xrFrag, we established stable

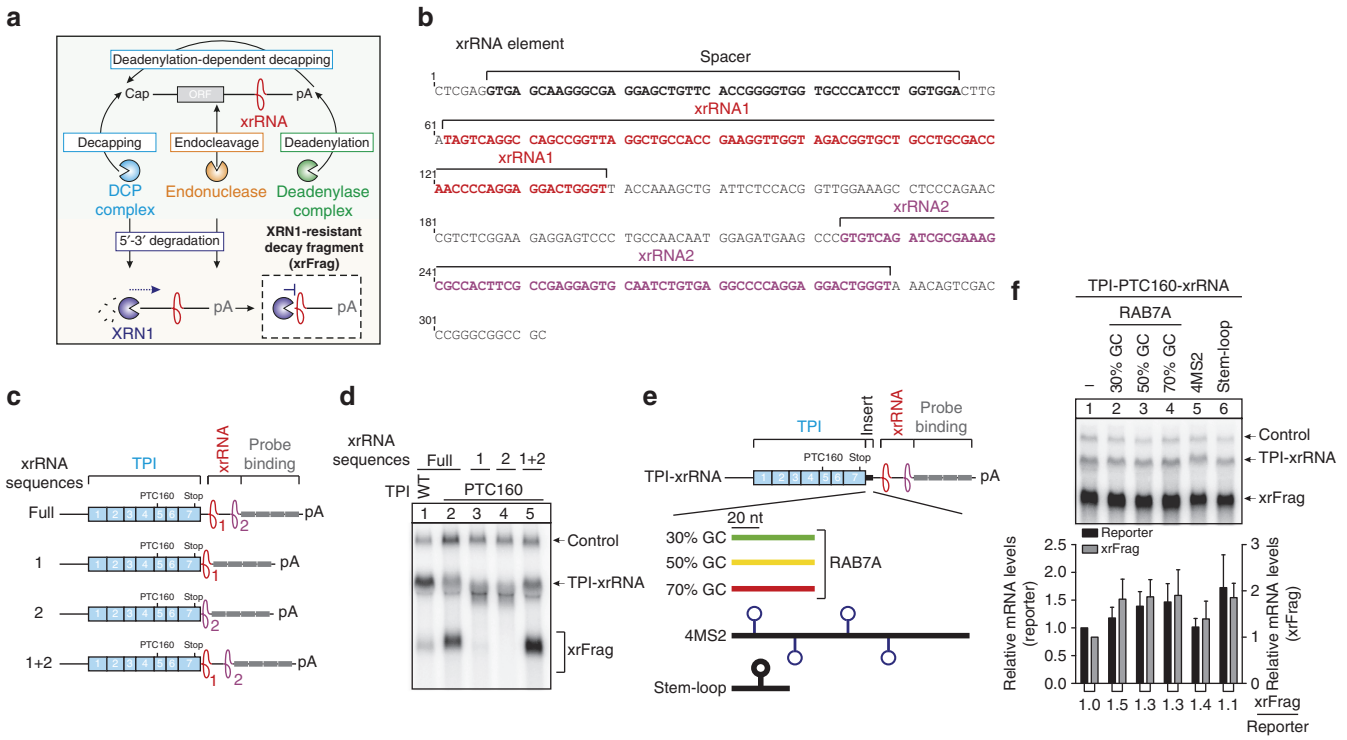


Figure 1 | Characterization of xrRNA elements enabling the detection of mRNA degradation intermediates. (**a**) Depicted are the general mRNA degradation pathways leading either directly (decapping and endocleavage) or indirectly (deadenylation) to 5′–3′ decay executed by XRN1. The presence of a stable XRN1-resistant RNA structure (xrRNA) prevents XRN1 from further progression and thus protects the remaining RNA fragment (xrFrag) from degradation from the 5′ end. (**b**) The DNA sequence of the xrRNA element used in reporter constructs is shown with annotations of sequence motifs. (**c**,**e**) Schematic representation of the TPI reporter mRNA. The TPI gene is depicted as blue boxes representing single exons (exon numbers indicated). The positions of the normal stop codons (stop) and premature translation termination codons are shown. Northern blot probe binding sites in the 3′ UTR are depicted as grey boxes and single xrRNA structures 1 and 2 are shown in red and purple. The difference between full xrRNA and 1 + 2 is the presence or absence of a short spacer region (indicated in **b**). (**e**) The 60 bp elements with varying GC content were derived from the RAB7A 3′ UTR, the 4MS2 binding sites are identical to those used in tethering experiments (Fig. 5) and the stem–loop structure is used in other reporters to block translation initiation (Fig. 5). (**d**,**f**) Northern blots of RNA samples extracted from HeLa cells transfected with the indicated reporter constructs. Co-transfected LacZ served as control mRNA. (**f**) Mean values of reporter and xrFrag signal ± s.d. ($n = 3$) were quantified and normalized to the TPI reporter without insert. The ratio of xrFrag to reporter mRNA levels is indicated below the graph.

tetracycline-inducible HeLa cell lines expressing β-globin or TPI reporter mRNAs with 3′ UTR xrRNA elements. Although we observed qualitatively similar results as with transient transfections, the overall effects were more pronounced when stable cell lines were analysed (Fig. 2d,e and Supplementary Fig. 1c,d). To determine the cellular degradation site of the PTC39 substrate, we isolated RNAs after sub-cellular fractionation. The full-length reporter was abundant in the nuclear fraction, but almost undetectable in the cytoplasm (Supplementary Fig. 1e). In contrast, xrFrags were preferentially found in the cytoplasm. These observations demonstrate that the xrFrags accumulate due to the degradation of the NMD substrate after the export to the cytosol.

We next examined the kinetics of the synthesis and decay of the reporter mRNAs by a time-course assay. Whereas the control WT mRNA increased in abundance during the first 8 h after induction, no PTC mRNA accumulated when UPF1 was present (Fig. 2f and Supplementary Fig. 1f). Interestingly, high xrFrag levels were detected already 4 h after induction. This indicates very rapid degradation kinetics of the PTC mRNAs, which was visualized via the concurrent detection of xrFrags. Inhibiting translation by cycloheximide treatment before induction abolished both, PTC reporter degradation and xrFrag generation (Supplementary Fig. 1g). Likewise, treating the cells with caffeine, a known inhibitor of the NMD-kinase SMG1 (ref. 27), resulted in a dose-dependent increase of PTC-reporter abundance accompanied by a decreased accumulation of xrFrag (Supplementary Fig. 1h). Taken together, the application of xrRNA elements enables the simultaneous detection of both full-length reporter mRNAs and 5′ processed decay intermediates, which in combination greatly improves the readout of RNA degradation analyses.

Next, we monitored the stability of the xrFrag as well as the TPI–WT and two different NMD substrate mRNAs (TPI-PTC160 and TPI-SMG5; Fig. 3a) for 6 h after inhibition of transcription by actinomycin D. Compared with TPI–WT the degradation of both NMD substrates was accelerated, albeit with different kinetics (Fig. 3b). In contrast, the xrFrags generated from the NMD reporter mRNAs were more stable and exhibited a delay before decay commenced, leading overall to longer apparent half-life time (Fig. 3b). To further study the turnover of xrFrags, we chased the decay intermediates of globin-PTC39 with actinomycin D treatment in XRN1-depleted cells after a 4 h induction of transcription (Fig. 3c). Interestingly, the readily visible 3′ fragments exhibited a similar decay kinetic as the xrFrags of control cells. By contrast, in XRN1-depleted cells the xrFrags increased in abundance during the first 4 h after transcriptional shutoff. This suggests that residual XRN1 molecules will continue to degrade 3′ fragments, leading to a constant formation of xrFrags from the pool of 3′ fragments. We

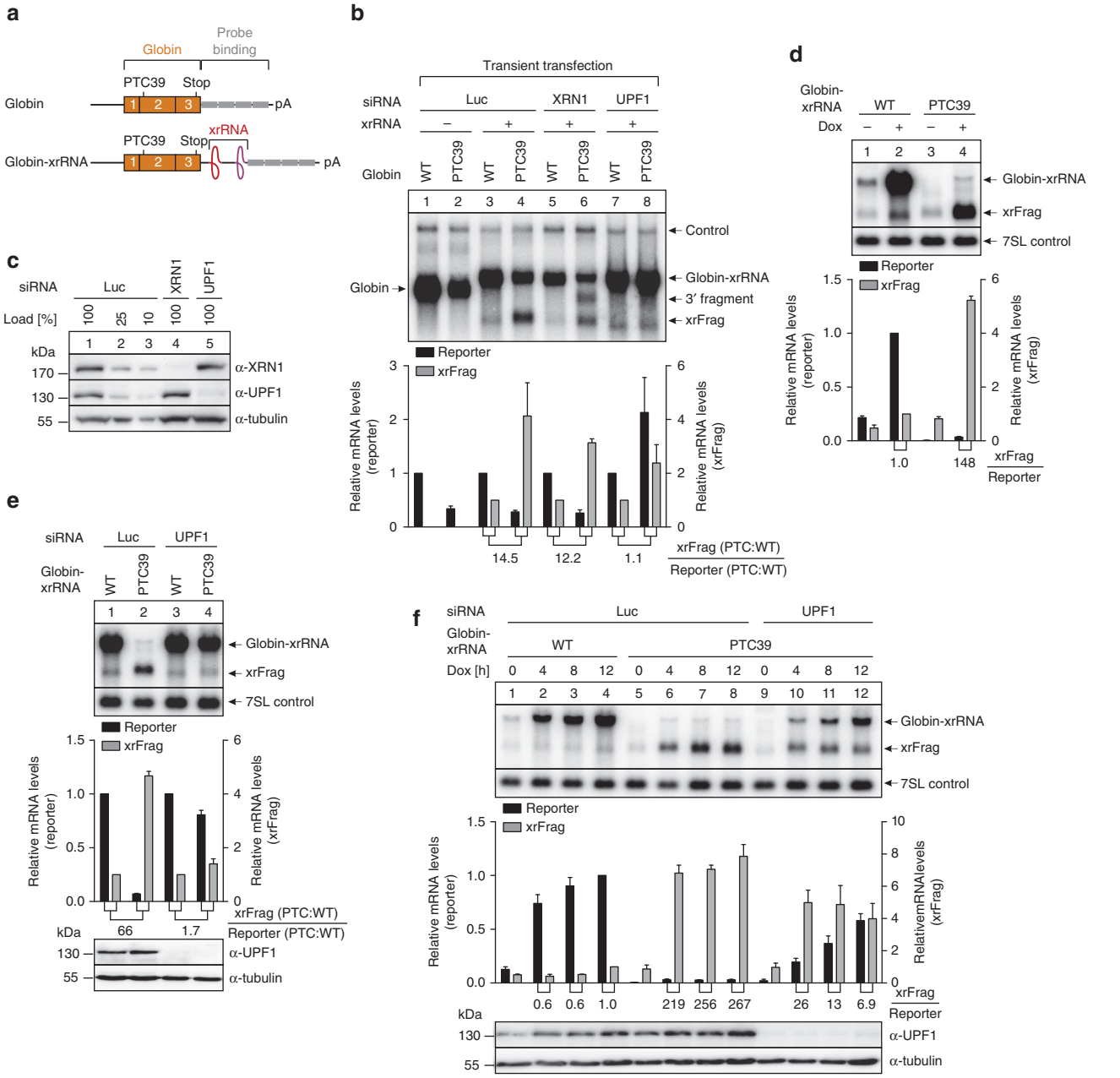


Figure 2 | Degradation of NMD substrates is traceable by xrFrag analysis. (**a**) Depiction of the β-globin reporter mRNAs as in Fig. 1. (**b**) Northern blot of RNA samples extracted from HeLa cells transfected with the indicated siRNAs and reporter constructs. Co-transfected LacZ served as control mRNA. Mean values of reporter and xrFrag signal ± s.d. ($n=3$) were quantified and for each knockdown condition the PTC values were normalized to the WT. The ratio of xrFrag to reporter mRNA levels is indicated below the graph. (**c**) Western blot analysis of siRNA knockdown efficiency using the indicated antibodies. Tubulin served as loading control. (**d**–**f**) Total RNA was extracted from stable HeLa Flp-In T-REx cells expressing the indicated reporter RNA and analysed by northern blotting. (**e**,**f**) The cells were transfected with the indicated siRNA 72 h before induction of expression. 7SL RNA served as endogenous control RNA. Unless indicated otherwise (**f**), reporter mRNA expression was induced for 24 h with 1 μg ml^{-1} doxycycline (Dox). Mean values of reporter and xrFrag signal ± s.d. ($n=3$) were quantified and normalized to the WT control (+ Dox for **d**; Luc or UPF1 knockdown for **e**; 12 h after Dox for **f**). The ratio of xrFrag to reporter mRNA levels is indicated below the graph.

therefore hypothesized that in XRN1-depleted cells the production of 3′ fragments would precede the accumulation of xrFrags. Indeed, while 3′ fragments were more abundant than xrFrags in the first hours after transcriptional induction, this ratio decreased after 8 h of continuous PTC reporter expression (Fig. 3d). To determine whether xrFrags are degraded over time via continuous 5′–3′ decay, we constructed mRNA reporters containing additional xrRNA elements (that is, in total 2x- and 3x-xrRNA elements; Supplementary Fig. 2a). Interestingly, expression of these reporters resulted in the appearance of additional xrFrags (Supplementary Fig. 2b,c). This finding suggests that in mammalian cells xrRNAs do not irreversibly trap, but impair XRN1 activity to such an extent that xrFrags can be detected.

Dissecting pathways involved in NMD using xrFrag analysis. In current models of NMD, phosphorylated and RNA-bound UPF1 serves as a docking platform for a variety of decay-inducing

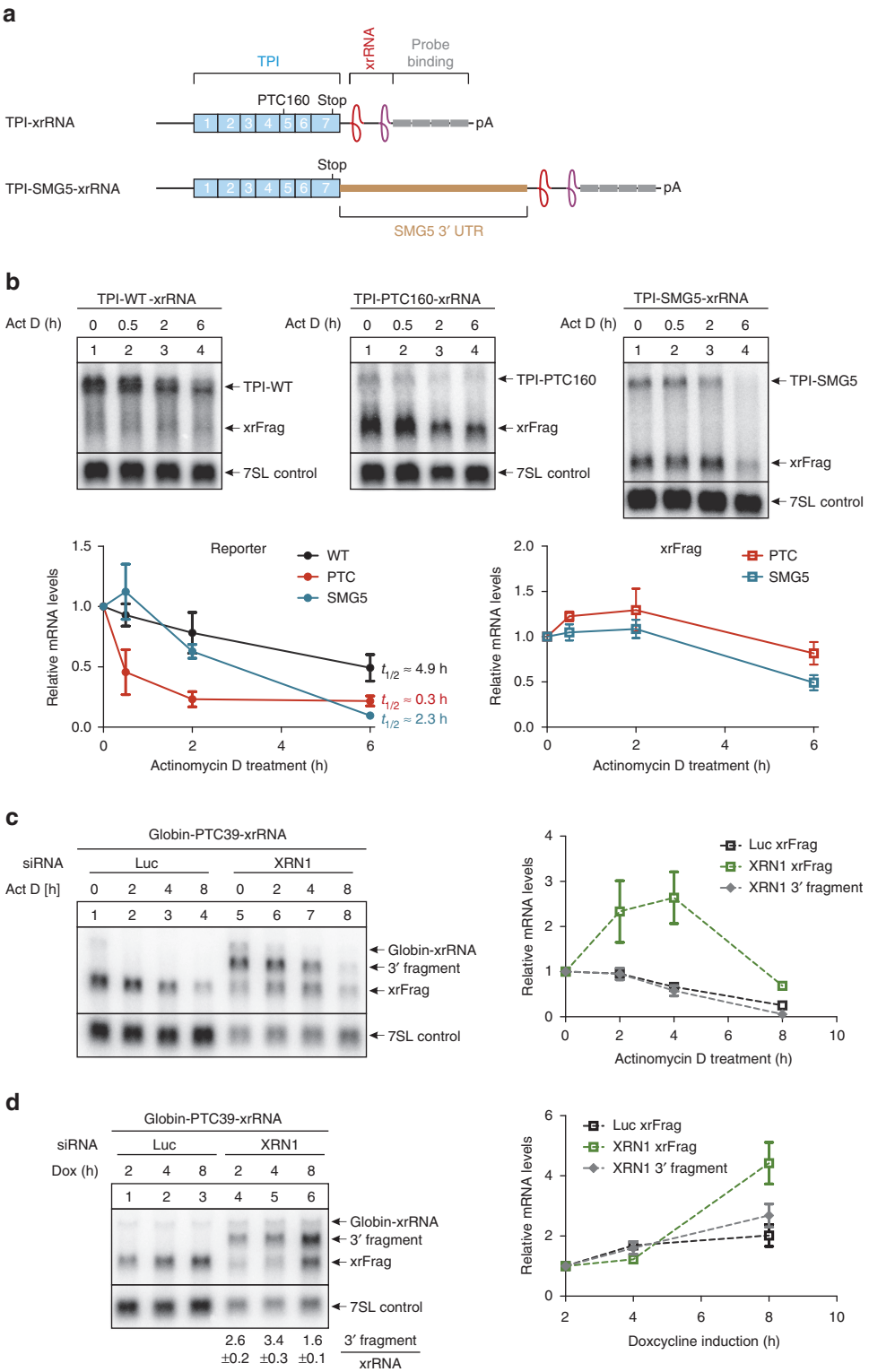


Figure 3 | Kinetics of xrFrag accumulation and degradation. (**a**) Schematic representation of the TPI reporter mRNA as in Fig. 1. The NMD-activating long 3′ UTR fragment of SMG5 is indicated. (**b**–**d**) Northern blot analysis of RNA samples derived from HeLa stable cell lines expressing the indicated reporter constructs as described in Fig. 2. Mean values of reporter, xrFrag and 3′ fragment signal ± s.d. ($n = 3$) were quantified and normalized to the 7SL endogenous control. (**b**) 2 h after induction of transcription by doxycycline (Dox), actinomycin D (5 μg ml^{-1}) was added and the cells harvested at the indicated time points. (**c**) Transcription was induced for 4 h with Dox and reporter mRNAs were chased for the indicated time with actinomycin D. (**d**) Induction of reporter transcription was performed for the indicated time with Dox.

factors (Fig. 4a). We decided to determine if the xrRNA system is able to detect differences between redundant mRNA decay pathways that are activated during NMD. To this end, we performed siRNA-mediated knockdowns of SMG6 and SMG7 (Fig. 4b). In line with published data[28], the depletion of either SMG6 or SMG7 did not completely stabilize PTC to WT mRNA

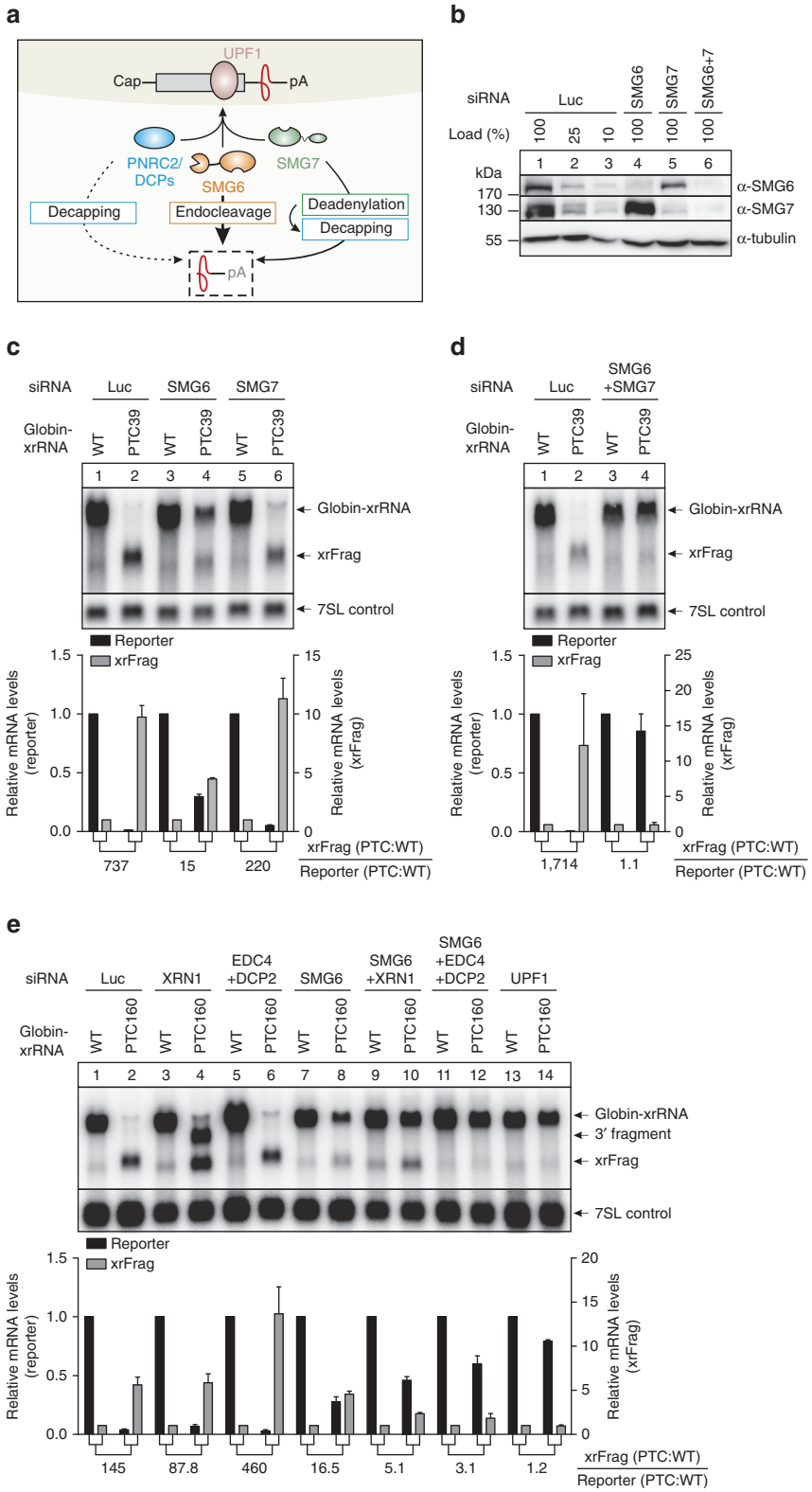


Figure 4 | Tracing distinct degradation pathways employed during NMD by monitoring xrFrags. (**a**) Model showing the recruitment of NMD-specific decay-inducing factors to activated UPF1 and their contribution to xrFrag accumulation. (**b**) Western blot analysis of knockdown efficiencies was performed with the indicated antibodies, tubulin served as loading control. (**c**–**e**) Northern blots of RNA samples extracted from stable HeLa cell lines transfected with the indicated siRNAs and expressing the indicated reporter constructs. Endogenous 7SL served as control RNA. Mean values of reporter and xrFrag signal ± s.d. ($n = 3$) were quantified and for each knockdown condition the PTC values were normalized to the WT. The ratio of xrFrag to reporter mRNA levels is indicated below the graph.

levels (Fig. 4c and Supplementary Fig. 3a). Nevertheless, we found that compared with SMG7 depletion, the knockdown of SMG6 decreased xrFrag:reporter ratios more profoundly for both β-globin and TPI reporter mRNAs. Furthermore, we detected almost identical xrFrag:reporter levels for WT and PTC reporter upon the combined SMG6–SMG7 knockdown, indicating near-complete NMD inhibition (Fig. 4d and Supplementary Fig. 3b). Taken together, these results are in line with the current NMD model, in which SMG6 is the major and SMG7 the alternative degradation-promoting factor.

We next aimed to distinguish decapping from endocleavage activities during NMD. To this end, we generated a reporter with a 5′ xrRNA (xrA). We inserted a 5′ stem–loop and a NanoLuc ORF upstream of the xrA to prevent ribosomes from disrupting the xrRNA structure. Downstream of the xrA, an EMCV IRES followed by the TPI ORF with or without a second xrRNA (xrB) was cloned (Supplementary Fig. 4a). The activation of NMD by IRES-mediated translation will either lead to endocleavage or decapping and the accumulation of xrFragB or xrFragA, respectively. We found decreased levels of the PTC-containing mRNA and increased xrFragB compared with the WT. Surprisingly, strongly reduced amounts of xrFragA were detected, too (Supplementary Fig. 4b). This finding underlines that as long as SMG6 is active in the NMD pathway, endocleavage dominates over decapping. Indeed, increased levels of PTC-containing xrFragA accumulated in SMG6-depleted cells, leading to identical xrFrag:reporter ratios of the PTC and the WT construct (Supplementary Fig. 4c).

If deadenylation-dependent decapping occurs during NMD, we expect that decapped full-length reporter mRNAs appear when the decapping machinery or XRN1 are depleted together with SMG6. Although we readily detect 3′ fragments as a result of endocleavage, knockdown of XRN1 alone lead to only marginally increased levels of the full-length reporter RNA (Fig. 4e and Supplementary Fig. 3c,d). To shut down the decapping machinery, we used siRNAs targeting the catalytic subunit DCP2 and the scaffold protein EDC4 (also called Hedls)[29]. We observed no accumulation of PTC reporter mRNA upon EDC4 + DCP2 depletion. When SMG6 and XRN1 were depleted together, no 3′ fragments were detected and reporter levels increased twofold compared with SMG6 knockdown alone, suggesting that 5′–3′ decay becomes more important when SMG6 is depleted. Surprisingly, also the combined SMG6 + EDC4 + DCP2 knockdown could not fully increase the PTC reporter to WT levels, even though elevated PTC reporter levels were observed. Interestingly, the knockdown of UPF1 increased the reporter RNAs more than the combined SMG6 + XRN1 or SMG6 + EDC4 + DCP2 depletion, suggesting that not all degradation during NMD can be attributed to decapping or endocleavage. However, we cannot exclude the possibility that the incomplete depletion of the decay factors contributes to the observed discrepancy. Taken together, xrFrag analysis suggests that 3′–5′ degradation pathways contribute to the removal of NMD substrates, when 5′–3′ decay is inhibited.

5′–3′ decay is initiated by decapping and endocleavage. Having investigated the decay routes of the complex NMD pathway in detail, we next used the MS2 tethering system to study individual mRNA-degrading proteins by recruiting them to an xrRNA-containing reporter mRNA (Fig. 5a).

Tethering of the DCP1 interacting and decapping-promoting protein PNRC2 reduced the levels of the TPI reporter mRNA and increased xrFrag levels more than twofold (Fig. 5b). The W114A mutant of PRNC2, which is unable to bind DCP1 (ref. 30), was inactive in the tethering assay. However, tethering of the ΔC mutant, which lacks the C-terminal NR box required for the UPF1 interaction[30], showed a degradation similar to the WT. In contrast to the decapping factor PNRC2, the NMD factor SMG7 recruits via its C-terminus the catalytic subunit POP2 of the CCR4-NOT complex leading to mRNA deadenylation[11]. The accelerated deadenylation is followed by decapping and degradation from the 5′ end[31]. Interestingly, when we tethered SMG7 full-length or only the C-terminal domain, which both lead to reduced TPI reporter mRNA, we detected slightly decreased rather than increased xrFrag levels (Fig. 5c). We therefore wanted to compare the impact of decapping, endocleavage or deadenylation on the generation of xrFrag. To this end we tethered the C-terminal domain of SMG7 (deadenylation), PNRC2 (decapping) and the EJC core component BTZ, which was shown to activate endocleavage via NMD[26]. Interestingly, tethering of PNRC2 and BTZ resulted in decreased reporter levels and increased levels of xrFrag, whereas SMG7 decreased the amounts of the reporter as well as xrFrag (Fig. 5d). The effects of tethered SMG7 and PNRC2 were also observed with a reporter (SL-NL-4MS2-xrRNA), which is inefficiently translated due to a strong stem–loop in the 5′ UTR (Fig. 5e–g). Hence, both proteins act in a translation-independent manner. Because NMD is a translation-dependent process, BTZ was inactive when tethered to the non-translatable SL-NL-4MS2-xrRNA mRNA. Our results suggest that SMG7-mediated degradation does not only enhance 5′–3′ decay, but also leads to the degradation of the substrate mRNA by 3′–5′ decay. This could also explain the remaining degradational activity we observed in the analysis of NMD substrates after combined knockdowns of SMG6 with XRN1 or decapping factors (Fig. 4e and Supplementary Fig. 3c). To test this idea, we constructed a tethering reporter containing a region from the MALAT1 lncRNA, which is cleaved by RNAse P and leads to the formation of a stable 3′ triple helix (Supplementary Fig. 5a). This structure was shown to stabilize mRNA, likely by preventing the 3′–5′ decay machinery to engage with the protected RNA 3′ end[32]. Tethering of PNRC2 leads to equal degradation of polyadenylated or triple-helix-containing RNA (tH), whereas SMG7 preferentially degraded the polyadenylated reporter mRNA (Supplementary Fig. 5b,c).

We next asked if the tethering of other deadenylation-promoting factors would lead to mRNA degradation with reduced xrFrag accumulation. For these experiments, we chose the silencing domains of TNRC6A and TNRC6B, which are involved in the miRNA-induced mRNA decay pathway, as well as TTP, which mediates the degradation of AU-rich elements (AREs)-containing mRNAs. These proteins were shown before to induce degradation by recruitment of deadenylase complexes, thereby leading to accelerated deadenylation[33–35]. In our tethering assay, we could confirm that all tested deadenylation-promoting proteins resulted in reporter degradation, however, no increase in xrFrag was detected for any of these factors (Fig. 5h). To further delineate the contribution of deadenylation to the observed decay, we made use of a reporter in which the xrRNA was located 5′ of the MS2 binding sites (Supplementary Fig. 5d). When SMG7-C, TNRC6A-SD or TNRC6B-SD were tethered, the resulting xrFrags showed a faster migration (Supplementary Fig. 5e), which was visible in lane profiles for the respective xrFrag signals (Supplementary Fig. 5f). Taken together, we find a differential enhancement of 5′–3′ degradation when comparing decapping-, endocleavage- and deadenylation-promoting factors. The xrRNA analysis therefore represents an ideal starting point to characterize whether a given protein induces 5′–3′ mRNA decay.

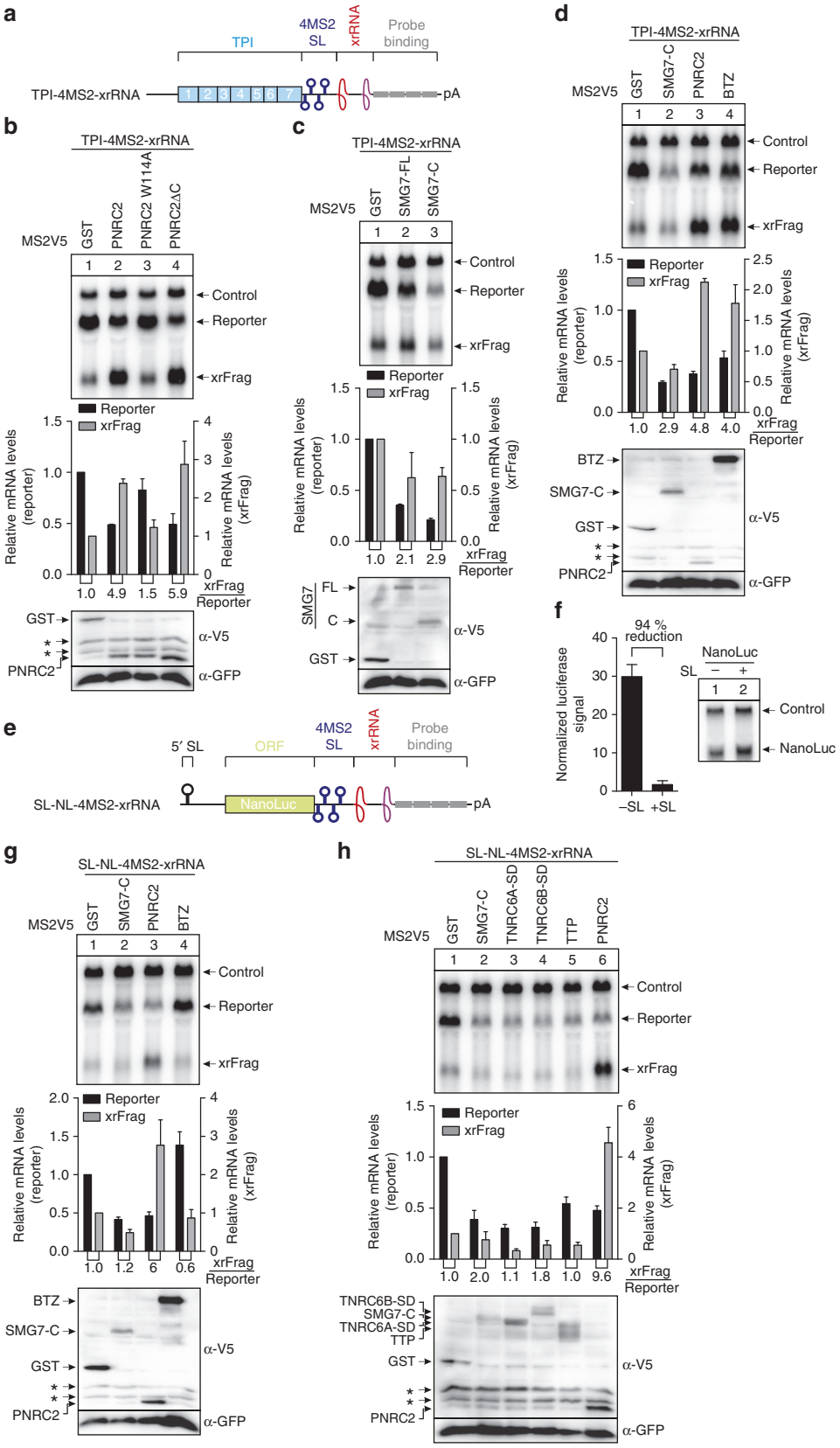


Figure 5 | Characterization of mRNA decay induced by direct protein tethering via xrFrag analysis. (**a**,**e**) TPI and NanoLuc tethering reporter mRNAs with 4MS2 binding sites are depicted as in Fig. 1. (**e**) The position of the 5′ stem–loop inhibiting ribosome scanning is indicated. (**b**–**d**, **g**,**h**) Northern blots of RNA samples extracted from HeLa cells transfected with the indicated tethering and reporter constructs. Mean values ± s.d. ($n=3$) for reporter and xrFrag levels were quantified and normalized to tethered GST, which served as control. The ratio of xrFrag to reporter mRNA levels is indicated below the graph. Western blots show the expression levels of the MS2V5-tagged constructs with GFP serving as transfection control. Unspecific bands are indicated with asterisks. (**f**) Translational efficiency (mean ± s.d., $n=3$) was measured by dual luciferase assay and compared for NanoLuc reporter with or without the 5′ stem–loop. Expression of the NanoLuc mRNAs is shown by northern blotting, co-transfected LacZ served as control.

Characterization of isolated deadenylation-inducing elements. So far, the xrFrag analysis allowed the dissection of complex decay pathways (NMD), as well as the characterization of degradation-inducing proteins (tethering). Next, we wanted to use xrRNAs to examine the turnover of mRNAs with individual decay elements, for example ARE and microRNA response elements (miRE). AREs are mRNA-destabilizing sequences that are known to work in heterologous reporter mRNAs[36]. In our constructs the c-fos or the granulocyte–macrophage colony-stimulating factor (GM–CSF) ARE were followed by a single xrRNA element (Fig. 6a). Upon induction of stable cell lines, we observed increased amounts of xrFrag for both AREs, but only the GM–CSF ARE also strongly reduced the expression of the reporter mRNA (Fig. 6b). However, calculating the xrFrag:reporter ratios indicated that both AREs enhanced 5′–3′ decay. The decay mediated by the GM–CSF ARE occurred via deadenylation-dependent decapping, because the reduced expression as well as increased xrFrag abundance was abolished by the depletion of either XRN1 or the CCR4-NOT scaffold protein CNOT1 (Fig. 6c,d).

The expression of many mammalian genes is regulated by miRNAs, which assemble an RNA-induced silencing complex onto their target mRNAs and thereby induce their degradation. We inserted into our reporter two previously described miRE (ref. 37) containing canonical let-7 and miR-21 binding sites (Fig. 6e). A reporter with mutated miREs was used as negative control. In comparison to the miRE (Mut) mRNA, the expression levels of the miRE (WT) mRNA upon induction of stable cell lines were strongly reduced (Fig. 6f). Interestingly, the xrFrag:reporter ratio was just increased by a factor of ~3, indicating that only some of the degradation occurred by 5′–3′ decay. Depletion of CNOT1 restored normal levels of reporter mRNA and xrFrag, confirming that miRE-mediated mRNA decay involves deadenylation by the CCR4-NOT complex (Fig. 6g). Our previous observation of deadenylation induced by TNRC6 tethering further supports this finding (Fig. 5h). In conclusion, xrRNA and xrFrag analysis can be used to study the characteristics of short mRNA-destabilizing sequence elements.

Endonucleolytic cleavage within TNF-α and IL6 3′ UTRs. Accelerated turnover of cytokine mRNAs represents a crucial regulation mechanism of immune responses, which is reported to utilize multiple decay elements and factors. To study this

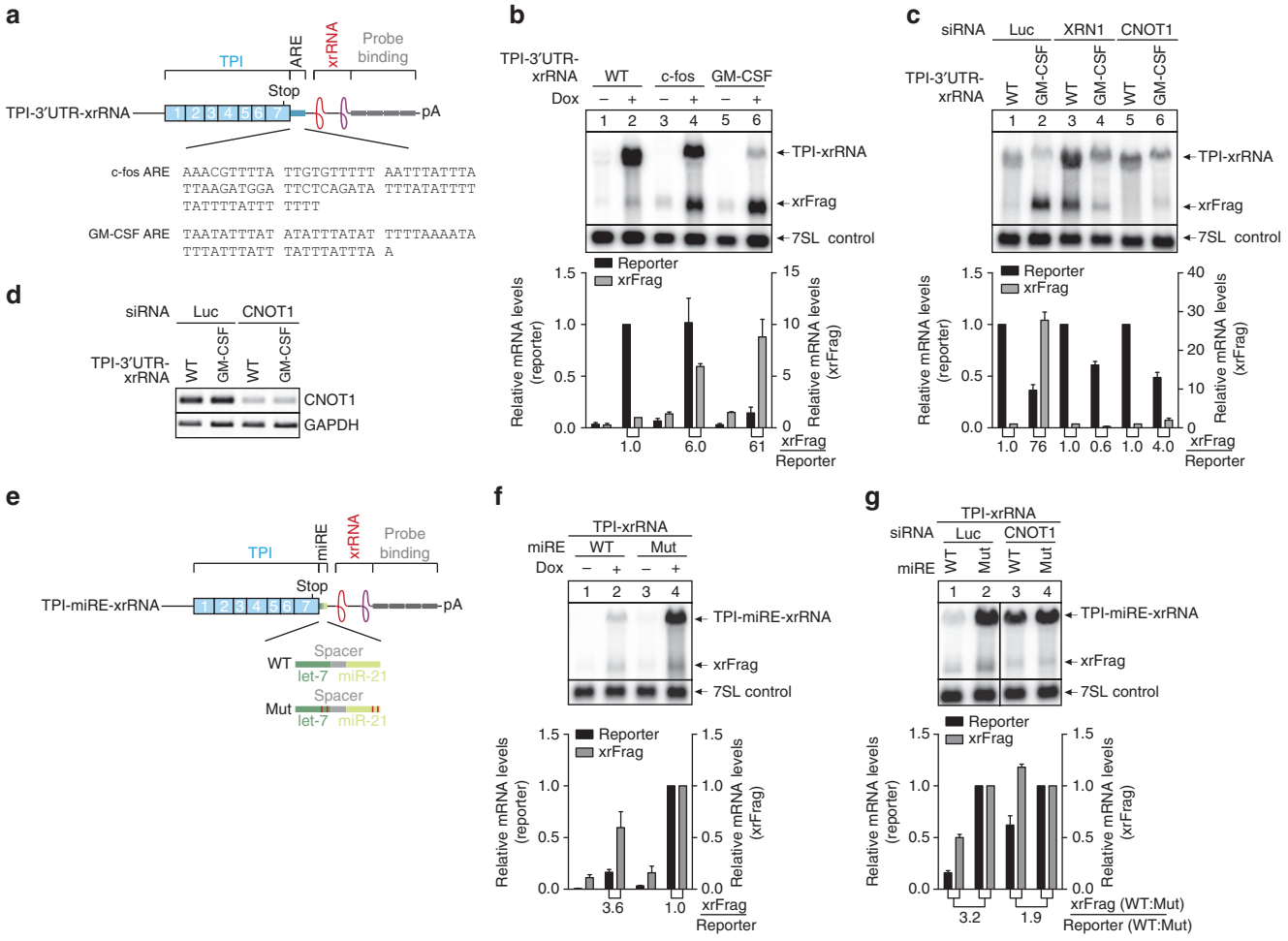


Figure 6 | Isolated ARE and miRNA decay elements induce deadenylation-dependent mRNA degradation. (**a**,**e**) Schematic representation of TPI reporter mRNAs depicted as in Fig. 1, containing (**a**) the decay-inducing AREs from c-fos and GM–CSF in the 3′ UTR or (**e**) WT or mutated miRE for let-7 and miR-21. (**b**,**c**, **f**,**g**) Hela Flp-In T-REx cells expressing the indicated constructs were harvested and RNA was extracted and analysed by northern blotting. (**c**,**g**) Knockdown was performed by transfecting the cells with the indicated siRNAs. Expression of reporter mRNAs was induced with 1 μg ml^{-1} doxycycline (+ Dox) for 24 h. Endogenous 7SL RNA levels are shown as a control. Mean values of reporter and xrFrag signal ± s.d. ($n = 3$) were quantified and normalized to the WT control (+ Dox for **b** and **f**; per knockdown condition for **c** and **g**). The ratio of xrFrag to reporter mRNA levels is indicated below the graph. (**d**) Semi-quantitative PCR analysis of the CNOT1 and GAPDH (control) expression levels in siRNA treated cells.

degradation pathway using our xrRNA system, we inserted the ARE- and conserved stem–loop-containing regions of the TNF-α and IL6 3′ UTRs between the TPI ORF and the xrRNA element (Fig. 7a and Supplementary Fig. 6a,b). As a negative control we used a similar sized fragment of the RAB7A 3′ UTR (ref. 24). For both reporters (TNF-α and IL6) we observed in transiently transfected cells a strong reduction of full-length mRNA levels, which was accompanied by an increase in xrFrag abundance (Fig. 7b). Hence, TNF-α and IL6 3′ UTRs induce mRNA degradation that involves 5′–3′ decay. In stable cell lines, the TNF-α and IL6 full-length transcripts were barely detectable 24 h after induction, whereas the xrFrag levels were strongly increased compared with the RAB7A control (Fig. 7c). In time-course assays the RAB7A-containing mRNA increased in expression for 8 h after induction. In contrast, the xrFrag generated from the degradation of the TNF-α and IL6 reporter mRNAs reached maximum levels as fast as 4 h after the addition of doxycycline (Fig. 7d). These fast decay kinetics are similar to those observed in PTC-containing TPI and globin mRNAs and suggest that TNF-α and IL6 reporter RNAs are very efficiently degraded. Treating the cells with the translation inhibitors cycloheximide or puromycin before induction caused a strong increase of both TNF-α- and IL6-containing reporter mRNAs and almost indistinguishable xrFrag:reporter ratios compared with the RAB7A control (Fig. 7e,f). Hence, the majority of the degradation involves translation-dependent mechanisms.

TNF-α as well as IL6 mRNAs have been reported to be degraded by Regnase-1-mediated endonucleolytic cleavage[4]. In addition, both 3′ UTRs contain several AREs, which can lead to deadenylation followed by decapping. We reasoned that the insertion of two xrRNA elements flanking the 3′ UTR of a reporter mRNA would allow to discriminate mRNAs undergoing decapping and endocleavage. To this end, we constructed reporter mRNAs with two xrRNA elements, one upstream (xrA) and one downstream (xrB) of the 3′ UTR of interest (Fig. 8a). Expression of the xrA-RAB7A-xrB control reporter leads to the production of both xrFrag species with roughly equal levels (Supplementary Fig. 7a). For the xrA-TNF-α-xrB reporter, the levels of the longer xrFragA were reduced, whereas the shorter xrFragB accumulated in a translation-dependent manner (Supplementary Fig. 7a,b). The production of xrFragB was very fast and occurred during the first 4 h after induction (Supplementary Fig. 7c).

If both deadenylation-dependent decapping and endocleavage decay routes are used for the degradation, decapped full-length transcripts as well as 3′ fragments are expected to accumulate in cells lacking XRN1. Surprisingly, we did not detect increased amounts of decapped, full-length TNF-α and IL6 reporters in XRN1-depleted cells (Fig. 8b,c). Instead, we observed additional bands representing 3′ fragments, indicating that the enhanced 5′–3′ decay involves endonucleolytic cleavage. A conserved stem–loop in the TNF-α and IL6 3′ UTRs has been reported as binding site of Regnase-1 (ref. 4). We therefore asked, if the observed 3′ fragments are generated by endocleavage close to the stem–loop sequence. To this end, we amplified, cloned and sequenced the 3′ fragments of IL6 and TNF-α. Surprisingly, the cleavage sites of both reporters were outside the stem–loop sequence, 1–20 nucleotides upstream of the IL6 stem–loop and 132–138 nucleotides downstream of the TNF-α stem–loop (Supplementary Fig. 6a,b). To further analyse the endonucleolytic cleavage of these transcripts, we generated constructs in which the observed cleavage sites were deleted (Fig. 8d). Interestingly, the deletion of the mapped cleavage site resulted in a weak (TNF-α) or moderate (IL6) increase of the reporter mRNAs and reduced levels of xrFrag (Fig. 8e). In contrast, the slightly lower expression levels (compared with RAB7A) of a reporter mRNA containing the isolated TNF-α cleavage site (Fig. 8d) resulted in mildly increased xrFrag accumulation (Fig. 8e). Upon XRN1 knockdown, TNF-α and IL6 3′ fragment levels were strongly reduced when the cleavage sites were deleted (Fig. 8f). Strikingly, in the TNF-α construct which contained the cleavage site, but not the stem–loop itself, we detected a 3′ endocleavage fragment. We mapped the cleavage site of this mRNA as described above to the same area as in the full-length TNF-α 3′ UTR (Supplementary Fig. 6a).

In conclusion, the xrFrag analysis helped to identify that TNF-α and IL6 3′ UTRs undergo endocleavage at previously uncharacterized positions outside the canonical Regnase-1 binding site. However, endocleavage can only partially explain the rapid degradation of both reporters. Hence, cytokine degradation also utilizes multiple major decay pathways for robust mRNA degradation, resembling the complex degradation mechanism observed for NMD substrates.

Discussion

The analysis of mRNA decay intermediates has greatly facilitated the understanding of mRNA turnover mechanisms. Previously, a poly(G) tract was used in yeast to inhibit exonucleases and to study the directionality of mRNA degradation pathways[38,39]. However, poly(G) tracts fail to trap mRNA turnover intermediates in mammalian cells. In this work, we establish a viral-derived XRN1-resistant RNA sequence (xrRNA) as a molecular tool to study mRNA decay in mammalian cells. The insertion of xrRNA into different reporter constructs results in the robust and reproducible accumulation of degradation intermediates (xrFrag). Since the xrFrag is produced during the decay of the xrRNA-bearing mRNA, it enables to unambiguously discriminate mRNA decay from other causes of reduced mRNA levels. Moreover, the use of xrRNA to visualize decay intermediates provides insights into the contribution of 5′–3′ decay pathways. Amongst its advantages are the minimal interference with other cellular processes, simplicity of its application and versatility. Although high levels of xrFrags could potentially inhibit cellular 5′–3′ decay due to the sequestration or inactivation of XRN1, as suggested previously[40] we did not observe signs of such an inhibition. In fact, we detect only a single xrFrag band on our northern blots, whereas the primary 3′ fragments were detected as an additional band exclusively in cells lacking XRN1.

Although the xrFrag method generates robust data in our hands, it also offers the possibility for future improvements. It will be interesting to explore if any xrRNA sequences can be identified or evolved that inhibit XRN1 to a greater extent than the MVE virus xrRNA used in this work. We provide evidence that the MVE-derived xrFrags represent meta-stable intermediates of 5′–3′ decay, which will be eventually degraded. It remains to be determined if more stable xrFrags could allow even deeper insights into the molecular mechanisms of mRNA degradation. Interestingly, for exact xrFrag analysis a single xrRNA is superior to a combination of two or three xrRNAs, which do not significantly elevate the amount of the 5′ xrFrag, but lead to the accumulation of additional xrFrags. This seemingly paradoxical observation is explained by the slow, but constant conversion of longer xrFrags into shorter xrFrags.

Many of the analyses presented in this work employ the xrFrag method to monitor the contribution of 5′–3′ decay pathways to mRNA turnover. However, it is conceivable that xrFrag analysis may also be used to discriminate deadenylation-dependent and -independent modes of decay. Due to the small size of the xrFrags, their running behaviour is highly dependent on their poly(A) tail lengths. In addition, the defined 5′ end of the xrFrags offer superior resolution compared with the variable 5′ ends observed

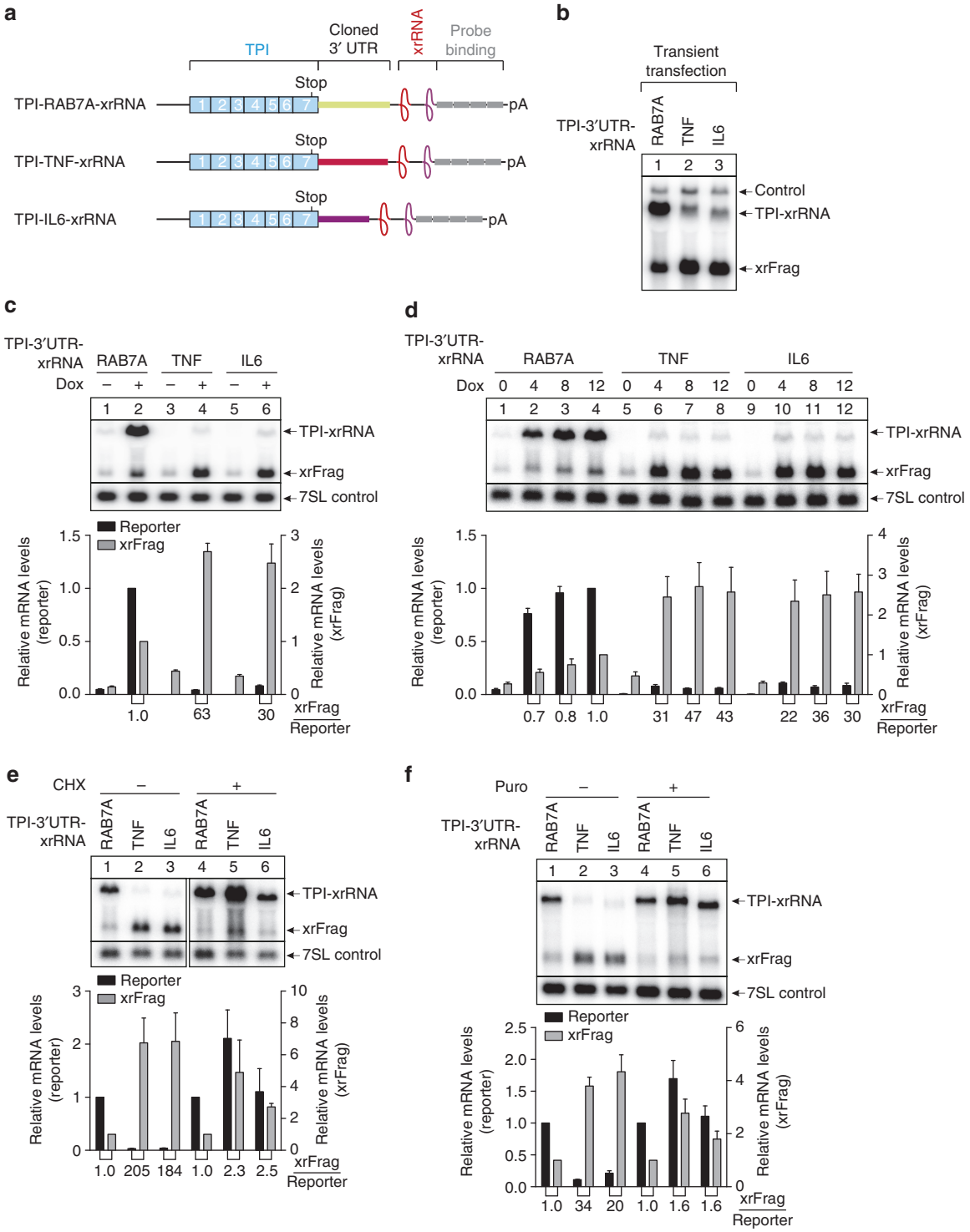


Figure 7 | The degradation induced by TNF-α and IL6 3′ UTRs is translation dependent. (**a**) TPI reporter mRNAs with RAB7A (control), TNF-α or IL6 (decay-inducing) 3′ UTRs are represented as in Fig. 1, with the inserted sequences shown as coloured boxes. (**b**) Northern blot analysis of RNA samples derived from HeLa cells transiently transfected with the indicated reporter constructs and LacZ as control. (**c**–**f**) HeLa Flp-In T-REx cells expressing the indicated reporter RNA were harvested, total RNA extracted and analysed by northern blotting. 7SL RNA serves as endogenous control RNA. Unless indicated otherwise (**c**,**d**), reporter mRNA expression was induced for 24 h with 1 μg ml^{-1} doxycycline (Dox). Cycloheximide treatment together with doxycycline induction was performed for 8 h with 100 μg ml^{-1} of cycloheximide. Puromycin (Puro) treatment together with doxycycline induction was performed for 4 h with 20 μg ml^{-1} of puromycin. Mean values of reporter and xrFrag signal ± s.d. ($n = 3$) were quantified and normalized to RAB7A control (+ Dox for **c**; 12 h after Dox for **d**; with or without cycloheximide for **e**; with or without Puro for **f**). The ratio of xrFrag to reporter mRNA levels is indicated below the graph.

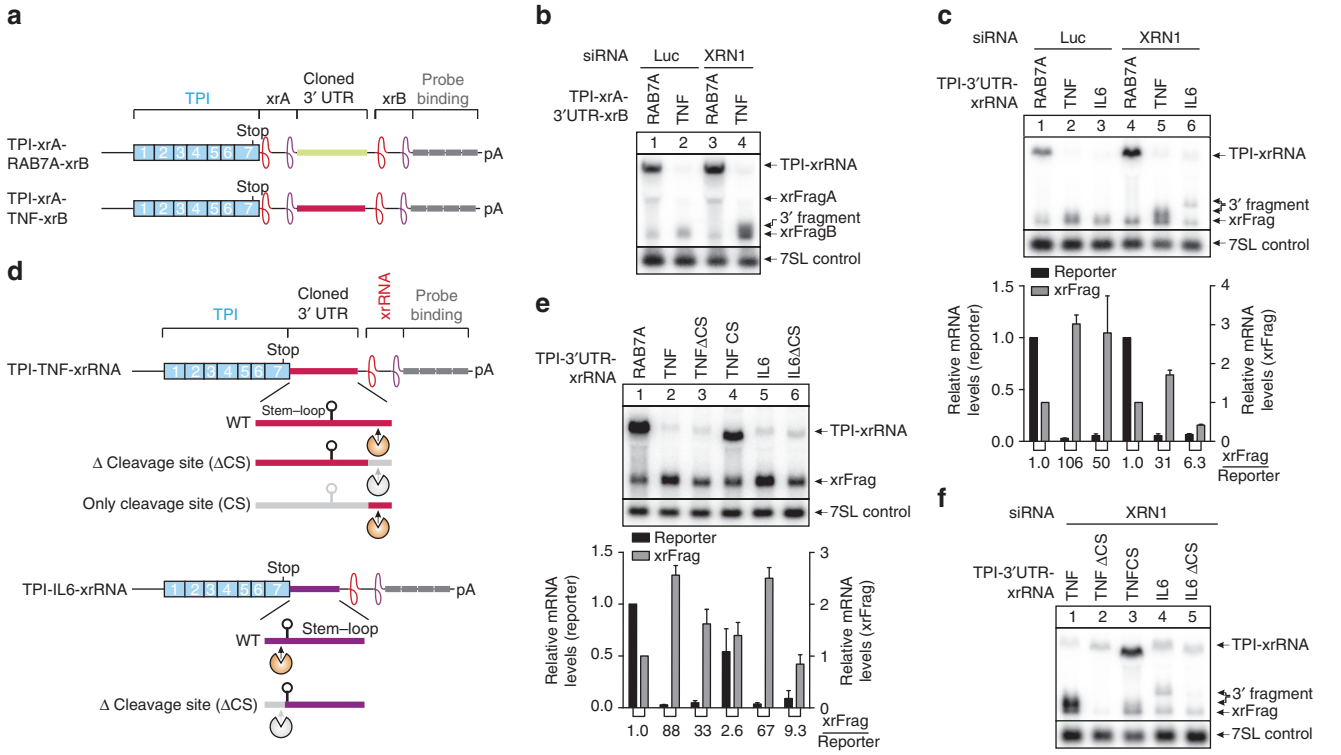


Figure 8 | The degradation induced by TNF-α and IL6 3′ UTRs involves endocleavage. (**a**,**d**) TPI reporter mRNAs with xrRNA-framed RAB7A (control), TNF-α or IL6 (decay-inducing) 3′ UTRs are represented as in Fig. 1, with the inserted sequences shown as coloured boxes. Dual xrRNA elements (xrA and xrB) are indicated. (**d**) Deletion mutants are shown containing the stem–loop, marked as a hairpin, or the cleavage site, marked by a schematic endonuclease. (**b**,**c**; **e**,**f**) HeLa Flp-In T-REx cells expressing the indicated reporter RNA were harvested, total RNA extracted and analysed by northern blotting. 7SL RNA serves as endogenous control RNA. The reporter mRNA expression was induced for 24 h with 1 μg ml^{-1} doxycycline (Dox). Knockdown was performed by transfecting the indicated siRNA. (**c**,**e**) Mean values of reporter and xrFrag signal ± s.d. ($n = 3$) were quantified and normalized to RAB7A control. The ratio of xrFrag to reporter mRNA levels is indicated below the graph.

for 3′ fragments that accumulate upon XRN1 depletion. Although some of our results indicate that xrFrags allow to assess poly(A) tail lengths, specific assays such as RNaseH-oligo(dT) digestion or PAT assays[41] are needed to improve this aspect of our experimental system.

The xrRNA realized its full potential when used in combination with other approaches and proved to be fully compatible with mRNA half-life measurements and the knockdown of specific mRNA decay factors. This allowed to dissect the contribution of 5′–3′ decay pathways to the overall degradation of an mRNA. Furthermore, assessing the pharmacological inhibition of mRNA decay is very robust with the xrRNA reporter as shown by our analysis of cycloheximide-, puromycin- and caffeine-treated cells. Finally, we were able to establish NMD reporters with a previously unprecedented dynamic range between WT and PTC-containing mRNAs. Standard NMD analyses with transiently transfected reporter constructs often show 5–10-fold differences between the WT and the PTC-containing steady-state mRNAs. Our system on the one hand displays a 10–50 fold reduction of the expression of the NMD substrate compared with its WT counterpart. In addition, the differential accumulation of xrFrag leads to a dynamic range of up to three orders of magnitude when the ratio of full-length reporter and xrFrag is calculated. Furthermore, the knockdown of the central NMD factor UPF1 increases the expression level of the PTC-containing mRNA to the level of the WT mRNA. To the best of our knowledge, this quantitatively complete inhibition of NMD has not been accomplished with any other assay to date. The performance of our NMD assay is even more remarkable in view of the possibility to follow the accumulation as well as the decay of NMD reporter and xrFrag over time. The degradation of the substrate mRNA occurs in a very short time frame and confirms that NMD eliminates bona fide substrate mRNAs with very rapid kinetics. In contrast, when XRN1 is abundant the xrFrags decay at a slower rate, which is similar to the one observed for 3′ fragments in XRN1-depleted cells. The comparable turnover rate of 3′ fragments and xrFrags is striking, given that their accumulation is caused by fundamentally different mechanisms of XRN1 inhibition. Importantly, we find that 3′ fragments are converted into xrFrags over time, which explains why xrFrags were still detected in XRN1-depleted cells.

The xrFrag analysis is also compatible with the tethering of RNA decay factors, which allows to disentangle complex mRNA turnover pathways, such as NMD. We show that the direct tethering of SMG7 to two different reporter mRNAs (translation-dependent and -independent) reduced the levels of the reporter mRNA, without substantial increase in the xrFrag:reporter ratio. The effect of tethered SMG7 was different from that of the tethered decapping activator PNRC2 or the tethered EJC-component BTZ, which both increased the amounts of xrFrags. Furthermore, SMG7 was less active in degrading a reporter mRNA containing the 3′ end of the MALAT1 lncRNA, which impairs 3′-5′ decay[32,42]. Our data are not only in line with the previously reported contribution of 3′–5′ decay to the overall activity of NMD[43], but also demonstrates the successful combination of the xrFrag analysis with well-established tethering assays.

Our analysis of AREs and miREs demonstrates that xrRNAs can also be used to study individual decay-inducing sequences. We find that both degradation elements require deadenylation activity by the CCR4-NOT1 complex, however only for AREs we observed a substantial increase in 5′–3′ decay. Combined with the results of tethered deadenylation-promoting factors, these findings support the notion that shortening of the poly(A) tails is not exclusively followed by enhanced decapping.

We also analysed mRNA turnover that is mediated by two well-studied cytokine 3′ UTRs, TNF-α and IL6. Different destabilizing sequences have been identified within these 3′ UTRs, for example, AREs. Furthermore, both 3′ UTRs contain a stem–loop structure, which is recognized by the RNA-decay factors Regnase-1 and Roquin. It has been suggested that Regnase-1 and Roquin degrade an overlapping set of mRNAs including inflammatory mRNAs (for example, TNF-α and IL6) via distinct mechanisms. Our experiments demonstrate that TNF-α and IL6 3′ UTRs are targeted by translation-dependent degradation including endonucleolytic cleavage in HeLa cells. We mapped the cleavage sites to positions up- and downstream of the conserved stem–loop sequences (Supplementary Fig. 6). Our observations for the IL6 3′ UTR are in good agreement with recently published results, although endocleavage did not occur at the putative binding site of Regnase-1 (ref. 4). However, the prominent endonucleolytic cleavage of the TNF-α 3′ UTR was unexpected in light of the preferential Roquin-mediated deadenylation that has been reported in the literature[4,44]. Furthermore, the site of endonucleolytic cleavage was clearly downstream of the reported Roquin and Regnase-1 binding sites. Deletion of the endocleavage sites partially impairs the degradation of TNF-α and IL6 3′ UTR reporter mRNAs, but also indicates that additional pathways are involved in the turnover of the reporters. Hence, xrFrag analysis confirms that multiple decay mechanisms contribute to the efficient mRNA degradation triggered by TNF-α and IL6 3′ UTRs.

In conclusion, we have established xrFrag analysis as a powerful tool to study mRNA turnover in mammalian cells and demonstrate its functionality by applying it to several mRNA decay pathways. Therefore, we envision that the use of xrRNAs will facilitate the analysis of mRNA decay under different experimental conditions and in various biological systems.

Methods

Plasmids and cell culture. Plasmid constructs β-globin WT and PTC39, TPI–WT and PTC160, TPI-4MS2 tethering reporter, LacZ control plasmid and the expression vector for MS2V5-BTZ were generated by inserting the respective DNA fragments[26,45] into the pCI-neo vector (Promega). Plasmids encoding for reporter constructs contained four copies of the northern blot probe binding site, as indicated in the figures. The MVE virus xrRNA element (nucleotides 10,488–10,725, reference sequence NC_000943.1, Fig. 1b) with a 50 nucleotide upstream linker was amplified from a gBlocks gene fragment (IDT) and inserted in the site indicated in the figures. Similarly, the 3′ UTR regions of RAB7A (nucleotides 862–1,311, reference sequence NM_004637.5), TNF-α (nucleotides 1,201–1,639, reference sequence NM_000594.3, Supplementary Fig. 6a) and IL6 (nucleotides 761–1,082, reference sequence NM_000600.4, Supplementary Fig. 6b), as well as the EMCV IRES and MALAT 3′ sequence were amplified from gBlocks gene fragments (IDT) or HeLa cDNA and cloned in the pCI vector. AREs (ref. 36) and miREs (ref. 37) were assembled from annealed primers. For stable cell lines, the reporters were cloned in the pcDNA5/FRT/TO vector. Full-length or mutants of SMG7, PNRC2, TNRC6A, TNRC6B and TTP were cloned from HeLa cDNA and inserted into the pCI-MS2V5 tethering vector. The NanoLuc ORF was cloned in pCI-neo containing a stable 5′ stem–loop (kindly provided by Gabriele Neu-Yilik), which was previously used to prevent cap-dependent translation in IRES-containing reporter constructs[46].

Transient transfections were done in HeLa Tet-Off cells (Clontech). Standard protocols were used to generate stable HeLa Flp-In T-REx cell lines (initially established by Elena Dobrikova and Matthias Gromeier, Duke University Medical Center). Expression of stable cell lines was induced with 1 μg ml^{-1} doxycycline for 24 h or for the indicated time before harvesting. All cell lines were cultured in DMEM (Gibco) supplemented with 9% foetal bovine serum (Gibco) and 1 × Pen Strep (Gibco) and the cells were incubated at 37 °C, 5% CO_2 and 90% humidity. Mycoplasma contamination was tested by PCR amplification of mycoplasma-specific genomic DNA[47].

siRNA transfections. For cells intended to be transiently transfected with plasmids, 5 × 10^5 HeLa Tet-Off cells were grown overnight in 6 cm plates and forward transfected with 300 pmol siRNA for single or 600 pmol total siRNA for double knockdowns using Lipofectamine RNAiMAX (Thermo Fisher)[26]. After 48 h the cells were counted and transferred to 6-well plates. Stable cell lines were reverse transfected using 2.5 μl Lipofectamine RNAiMAX and 60 pmol siRNA per 2 × 10^5 HeLa cells. The following siRNA target sequences were used: luciferase (5′-CGT ACGCGGAATACTTCGA-3′), XRN1 (5′-AGATGAACTTACCGTAGAA-3′), UPF1 (5′-GATGCAGTTCCGCTCCATT-3′), SMG6 (5′-GGGTCACAGTGC TGAAGTA-3′), SMG7 (5′-CGATTTGGAATACGCTTTA-3′), CNOT1 (5′-GGA ACUUGUUUGAAGAAUA-3′), DCP2 (5′-GGACTGGCTTTCTCGAAGA-3′), for EDC4 an equal mix of EDC4-1 (5′-GAGTTAAAGATGTGGTGTA-3′) and EDC4-2 (5′-TACACCACATCTTTAACTC-3′). The efficiency of the CNOT1 knockdown was confirmed by semi-qPCR using the following primers: 5′-AATGTTGGCCTG TCTGCAAG-3′, 5′-TGTCATTCCAGCAAGAGGGT-3′.

Transient plasmid transfections. 2.8 × 10^5 HeLa Tet-Off cells were seeded in 6-well plates 24 h before transfection by calcium phosphate precipitation with 0.5 μg of a mVenus expression plasmid, 2.5 μg control plasmid (LacZ) and 2 μg plasmid encoding for reporter mRNA. For tethering assays, 1.5 μg reporter, 1.5 μg of MS2V5-tagged expression plasmid and 3 μg control plasmid (LacZ) were transfected.

RNA extraction and northern blotting. Total RNA was extracted from individual replicates with peqGOLD TriFast (Peqlab), resolved on a 1% agarose and 0.4 M formaldehyde gel using the tricine-triethanolamine buffer system[48] and analysed by northern blotting. PSP65-globin plasmid was linearized with BamHI and used as template for *in vitro* transcribed [α-^{32}P]-GTP body-labelled RNA probes, which were used for the detection of all reporter and LacZ control RNA. 7SL endogenous RNA was detected using a 5′-^{32}P-labelled oligonucleotide (5′-TGCTCCGTTTCC GACCTGGGCCGGTTCACCCCTCCTT-3′). Signals were scanned using a Typhoon FLA 7000 (GE Healthcare) and raw unmodified scans were quantified using ImageQuant TL (GE Healthcare). Representative blots of at least three replicates are shown. Blots have been cropped and contrast-adjusted for presentation purposes. Uncropped images of main figures are collectively shown in Supplementary Fig. 8. As an example, positions of RNA marker bands are shown in Supplementary Figs 2 and 5. For sub-cellular fractionation, the cells were first harvested in polysome buffer (10 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris-HCl pH 7.4, 1% Triton X-100, 1% Na-deoxycholate and 1 mM DTT). Cytoplasmic and nuclear fractions were separated by centrifugation and RNA subsequently extracted from each fraction using peqGOLD TriFast. For mRNA half-life calculations, the one-phase-decay equation from GraphPad Prism 5 was used.

Immunoblot analysis and antibodies. SDS–polyacrylamide gel electrophoresis and immunoblot analysis were performed using protein samples derived from peqGOLD TriFast extractions or parallel transfection harvested with RIPA buffer.

The antibody against tubulin (T6074; 1:3,000 dilution) was from Sigma, the antibody against V5 (18870; 1:3,000 dilution) was from QED Bioscience, the antibodies against GFP (ab290; 1:3,000 dilution) and SMG6 (ab87539; 1:1,000 dilution) were from Abcam, the antibodies against XRN1 (A300-443A; 1:2,000 dilution), SMG7 (A302-170 A; 1:2,000 dilution) and EDC4 (A300-745A; 1:1,000) were from Bethyl and the antibody against UPF1 was kindly provided by Jens Lykke-Andersen (1:1,000 dilution). Secondary horseradish peroxidase-coupled antibodies against rabbit (111-035-006; 1:3,000 dilution) or mouse (115-035-003; 1:3,000 dilution) were from Jackson ImmunoResearch. Western Lightning Plus-ECL Enhanced Chemiluminescence Substrate (PerkinElmer) in combination with the myECL Imager (ThermoFisher) was used for visualization. Blots have been cropped and contrast-adjusted for presentation purposes. Uncropped images of main figures including molecular weight markers are collectively shown in Supplementary Fig. 8.

3′ fragment cloning. Total RNA (25 μg) was used for linker ligation using T4 RNA ligase I and 200 pmol of barcoded RNA linker (rGrCrUrGrArUrGrGrC rGrArUrGrArArUrGrArNrNrNrNrNrArArA)[26]. After RT–PCR with VNN-oligo(dT)$_{20}$ primer, TNF-α- or IL6-specific primers together with an adaptor-specific primer (Adaptor: 5′-GCTGATGGCGATCAATGA-3′; TNF-α: 5′-TTT TTTTGCGGCCGCCCGGGTCGACTAGTAGGGCGATTACAGACACA-3′; IL6: 5′-TTTTTTGCGGCCGCCCGGGTCGACTGAGGTAAGCCTACACTT TC-3′) were used for PCR amplification with Accuprime Taq Polymerase (Life Technologies). Amplified fragments were subsequently cloned in pGEM-T vector, sequenced by GATC Biotech and unique clones were used for mapping analysis.

Data Availability. The authors declare that all the data supporting the findings of this study are available within the article and its Supplementary Information files and from the corresponding authors upon reasonable request.

References

1. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455,** 58–63 (2008).
2. Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mRNA decay. *Nat. Rev. Genet.* **13,** 246–259 (2012).
3. Schott, J. & Stoecklin, G. Networks controlling mRNA decay in the immune system. *Wiley Interdiscip. Rev. RNA* **1,** 432–456 (2010).
4. Mino, T. *et al.* Regnase-1 and roquin regulate a common element in inflammatory mRNAs by spatiotemporally distinct mechanisms. *Cell* **161,** 1058–1073 (2015).
5. Anderson, P. Post-transcriptional control of cytokine production. *Nat. Immunol.* **9,** 353–359 (2008).
6. Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* **16,** 665–677 (2015).
7. Fatscher, T., Boehm, V. & Gehring, N. H. Mechanism, factors, and physiological role of nonsense-mediated mRNA decay. *Cell. Mol. Life Sci.* **72,** 4523–4544 (2015).
8. Mitchell, S. F. & Parker, R. Principles and properties of eukaryotic mRNPs. *Mol. Cell* **54,** 547–558 (2014).
9. Singh, G., Pratt, G., Yeo, G. W. & Moore, M. J. The clothes make the mRNA: past and present trends in mRNP fashion. *Annu. Rev. Biochem.* **84,** 325–354 (2015).
10. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* **8,** 113–126 (2007).
11. Loh, B., Jonas, S. & Izaurralde, E. The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev.* **27,** 2125–2138 (2013).
12. Tharun, S. Lsm1-7-Pat1 complex: a link between 3' and 5'-ends in mRNA decay? *RNA Biol.* **6,** 228–232 (2009).
13. Nishimura, T. *et al.* The eIF4E-binding protein 4E-T is a component of the mRNA decay machinery that bridges the 5' and 3' termini of target mRNAs. *Cell Rep.* **11,** 1425–1436 (2015).
14. Jinek, M., Coyle, S. M. & Doudna, J. A. Coupled 5' nucleotide recognition and processivity in Xrn1-mediated mRNA decay. *Mol. Cell* **41,** 600–608 (2011).
15. Muhlrad, D., Decker, C. J. & Parker, R. Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. *Genes Dev.* **8,** 855–866 (1994).
16. Tomecki, R. *et al.* The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. *EMBO J.* **29,** 2342–2357 (2010).
17. Lubas, M. *et al.* Exonuclease hDIS3L2 specifies an exosome-independent 3'-5' degradation pathway of human cytoplasmic mRNA. *EMBO J.* **32,** 1855–1868 (2013).
18. Viegas, S. C., Silva, I. J., Apura, P., Matos, R. G. & Arraiano, C. M. Surprises in the 3'-end: 'U' can decide too! *FEBS J.* **282,** 3489–3499 (2015).
19. Lim, J. *et al.* Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* **159,** 1365–1376 (2014).
20. Schoenberg, D. R. Mechanisms of endonuclease-mediated mRNA decay. *Wiley Interdiscip. Rev. RNA* **2,** 582–600 (2011).
21. Eberle, A. B., Lykke-Andersen, S., Muhlemann, O. & Jensen, T. H. SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat. Struct. Mol. Biol.* **16,** 49–55 (2009).
22. Braun, J. E. *et al.* A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. *Nat. Struct. Mol. Biol.* **19,** 1324–1331 (2012).
23. Chapman, E. G. *et al.* The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science* **344,** 307–310 (2014).
24. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited. *Trends Genet.* **29,** 569–574 (2013).
25. Lykke-Andersen, S. *et al.* Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.* **28,** 2498–2517 (2014).
26. Boehm, V., Haberman, N., Ottens, F., Ule, J. & Gehring, N. H. 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep.* **9,** 555–568 (2014).
27. Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y. & Ohno, S. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.* **15,** 2215–2228 (2001).
28. Jonas, S., Weichenrieder, O. & Izaurralde, E. An unusual arrangement of two 14-3-3-like domains in the SMG5-SMG7 heterodimer is required for efficient nonsense-mediated mRNA decay. *Genes Dev.* **27,** 211–225 (2013).
29. Erickson, S. L. *et al.* Competition between decapping complex formation and ubiquitin-mediated proteasomal degradation controls human Dcp2 decapping activity. *Mol. Cell. Biol.* **35,** 2144–2153 (2015).
30. Lai, T. *et al.* Structural basis of the PNRC2-mediated link between mrna surveillance and decapping. *Structure* **20,** 2025–2037 (2012).
31. Chen, C. Y. & Shyu, A. B. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA* **2,** 167–183 (2011).
32. Brown, J. A. *et al.* Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.* **21,** 633–640 (2014).
33. Lazzaretti, D., Tournier, I. & Izaurralde, E. The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins. *RNA* **15,** 1059–1066 (2009).
34. Braun, J. E., Huntzinger, E., Fauser, M. & Izaurralde, E. GW182 proteins directly recruit cytoplasmic deadenylase complexes to miRNA targets. *Mol. Cell* **44,** 120–133 (2011).
35. Fabian, M. R. *et al.* Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. *Nat. Struct. Mol. Biol.* **20,** 735–739 (2013).
36. Chen, C. Y., Xu, N. & Shyu, A. B. mRNA decay mediated by two distinct AU-rich elements from c-fos and granulocyte-macrophage colony-stimulating factor transcripts: different deadenylation kinetics and uncoupling from translation. *Mol. Cell. Biol.* **15,** 5777–5788 (1995).
37. Zhao, Y. *et al.* MicroRNA-mediated repression of nonsense mRNAs. *Elife* **3,** e03032 (2014).
38. Muhlrad, D. & Parker, R. Premature translational termination triggers mRNA decapping. *Nature* **370,** 578–581 (1994).
39. Vreken, P. & Raue, H. A. The rate-limiting step in yeast PGK1 mRNA degradation is an endonucleolytic cleavage in the 3'-terminal part of the coding region. *Mol. Cell. Biol.* **12,** 2986–2996 (1992).
40. Moon, S. L. *et al.* A noncoding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability. *RNA* **18,** 2029–2040 (2012).
41. Murray, E. L. & Schoenberg, D. R. Assays for determining poly(A) tail length and the polarity of mRNA decay in mammalian cells. *Methods Enzymol.* **448,** 483–504 (2008).
42. Akiyama, B. M., Eiler, D. & Kieft, J. S. Structured RNAs that evade or confound exonucleases: function follows form. *Curr. Opin. Struct. Biol.* **36,** 40–47 (2016).
43. Lejeune, F., Li, X. & Maquat, L. E. Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities. *Mol. Cell* **12,** 675–687 (2003).
44. Leppek, K. *et al.* Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* **153,** 869–881 (2013).
45. Steckelberg, A. L., Boehm, V., Gromadzka, A. M. & Gehring, N. H. CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep.* **2,** 454–461 (2012).
46. Holbrook, J. A., Neu-Yilik, G., Gehring, N. H., Kulozik, A. E. & Hentze, M. W. Internal ribosome entry sequence-mediated translation initiation triggers nonsense-mediated decay. *EMBO Rep.* **7,** 722–726 (2006).
47. Young, L., Sung, J., Stacey, G. & Masters, J. R. Detection of Mycoplasma in cell cultures. *Nat. Protoc.* **5,** 929–934 (2010).
48. Mansour, F. H. & Pestov, D. G. Separation of long RNA by agarose-formaldehyde gel electrophoresis. *Anal. Biochem.* **441,** 18–20 (2013).

Acknowledgements

We thank Heidi Thelen and Juliane Hancke for technical assistance; the Leptin lab for sharing equipment; Valentina Potrich for sharing protocols; Gabriele Neu-Yilik for providing the pCI-neo stem–loop vector; Elena Dobrikova and Matthias Gromeier for establishing and Matthias Hentze for sharing the HeLa Flp-In T-REx cells; Anna-Lena Steckelberg for critical reading of the manuscript and members of the Gehring lab for useful discussions. We are grateful to Jens Lykke-Andersen for the antibody against UPF1. This research was funded by a grant from the Deutsche Forschungsgemeinschaft (GE2014/4-1) to N.H.G.

Author contributions

N.H.G and V.B. conceived the study. V.B., J.V.G., M.C.M. and N.H.G. performed the experiments. V.B., J.V.G. and N.H.G. wrote the manuscript. All authors discussed the results and edited the manuscript.

Additional information

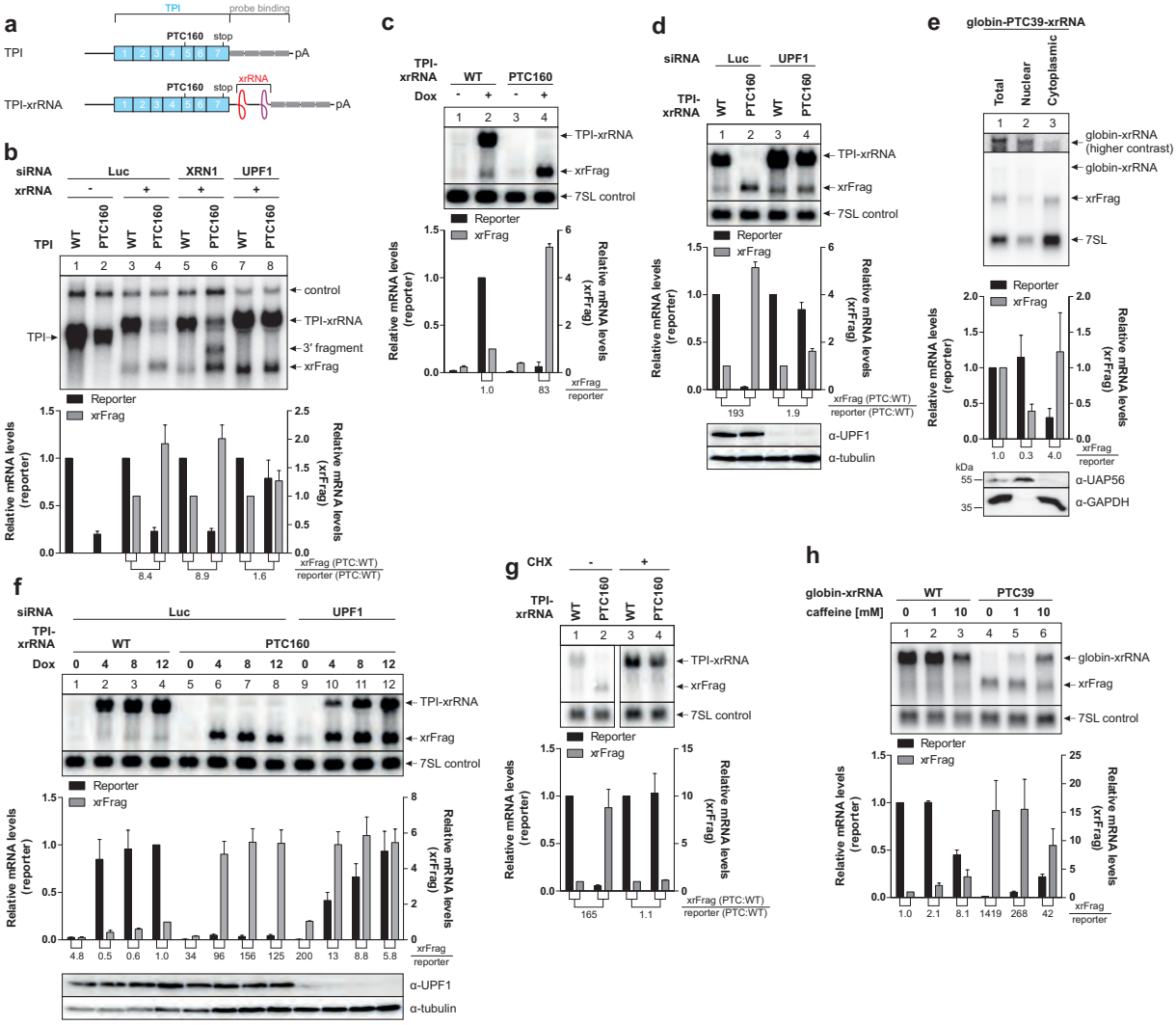
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Boehm, V. *et al.* Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences. *Nat. Commun.* **7**, 13691 doi: 10.1038/ncomms13691 (2016).

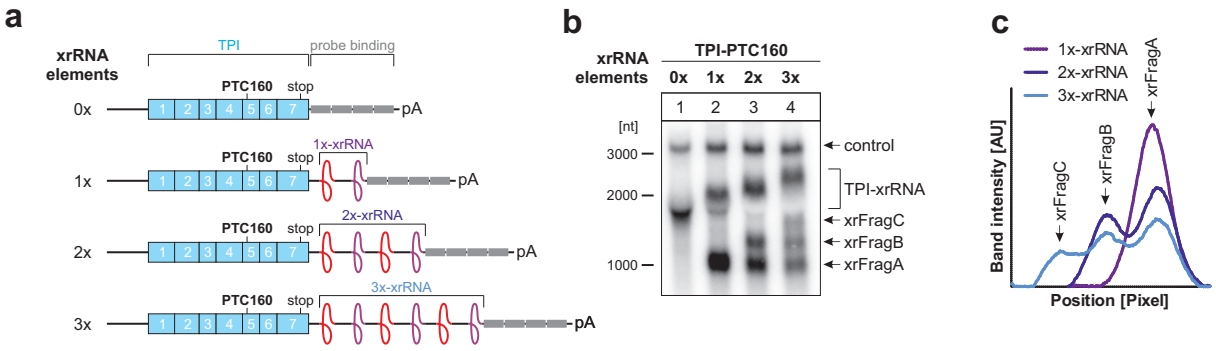
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Supplementary Figure 1

NMD-dependent accumulation of xrFrags derived from reporter mRNA degradation

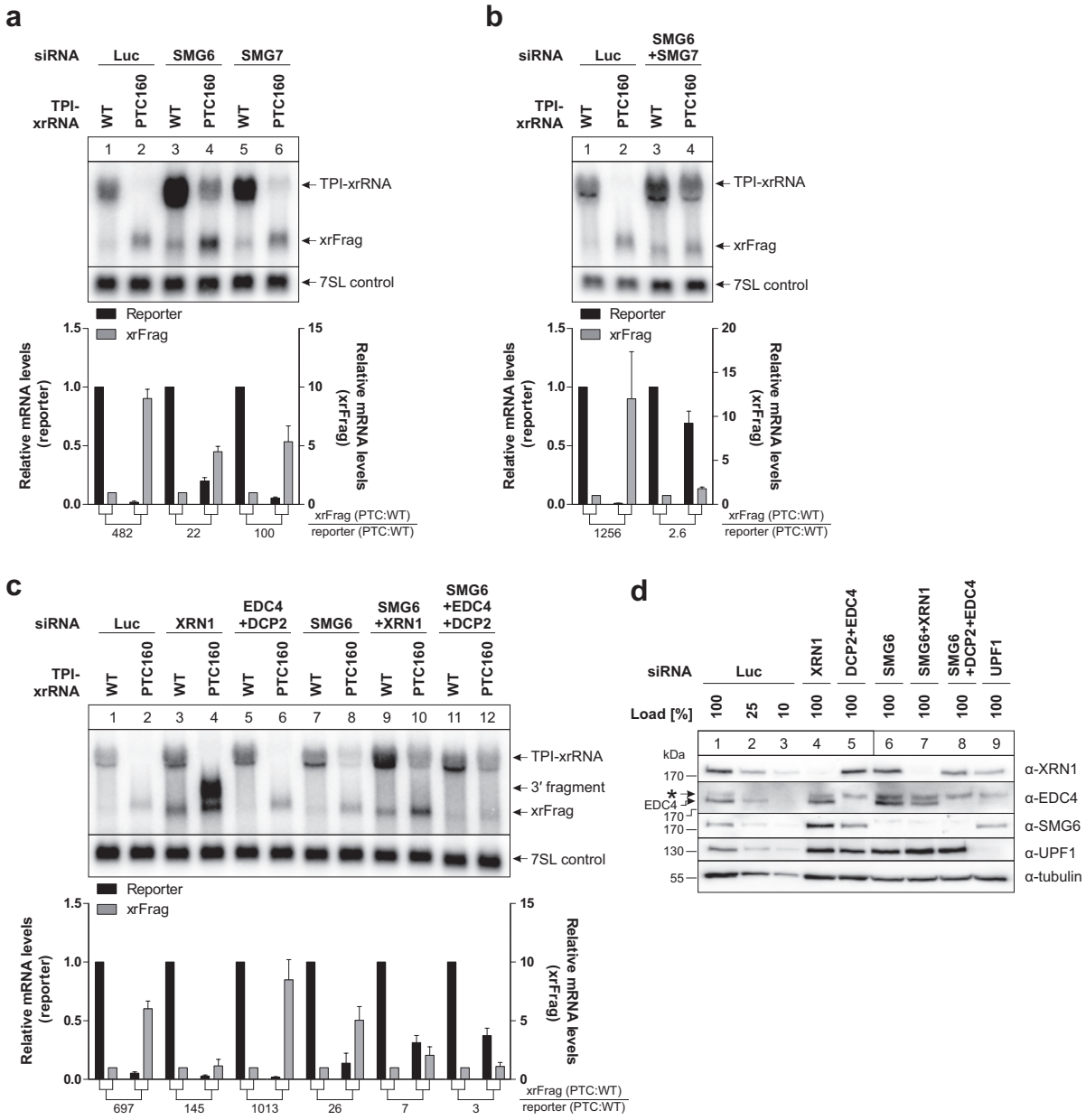
(a) WT or PTC-containing TPI reporter mRNAs with or without xrRNA are depicted as in Fig. 1. (b-h) Northern blots of total RNA extracted from transiently transfected HeLa cells (b) or stable cell lines (c-h) expressing the indicated reporter mRNAs. Co-transfected LacZ or endogenous 7SL served as control RNA. Mean values of reporter and xrFrag signal ± s.d. (n = 3) were quantified and for each knockdown or treatment condition the PTC values were normalized to the WT. For sub-cellular fractionation (e), the values for reporter and xrFrag mRNA levels were normalized to the respective value of the total RNA fraction. The ratio of xrFrag to reporter mRNA levels is indicated below all graphs. Successful fractionation was confirmed by the nuclear marker UAP56 and the cytoplasmic marker GAPDH (d, f) Western blot analysis of total protein extracted from the same samples with UPF1 antibody, tubulin served as loading control.



Supplementary Figure 2

Degradation of xrFrags via 5′-3′ decay

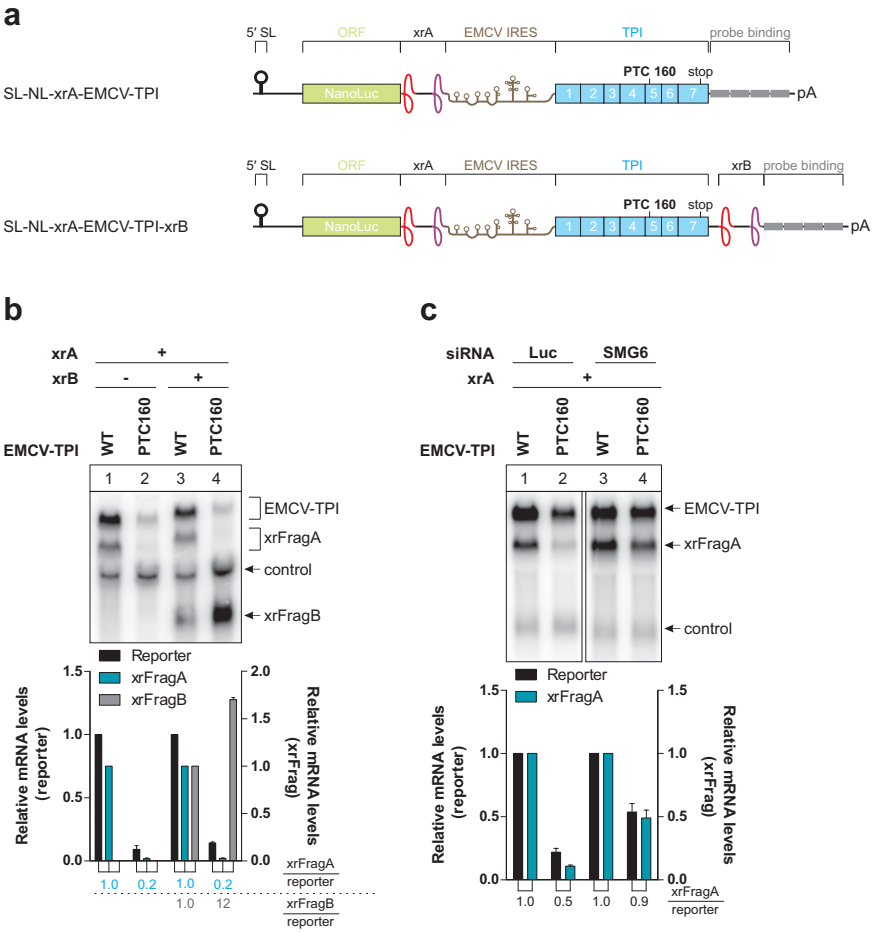
(a) Schematic representation of the PTC-containing TPI reporter mRNAs as in Fig. 1, indicating the multiple xrRNA element repeats. (b) Northern blot of total RNA extracted from transiently transfected HeLa cells expressing the indicated reporter mRNAs. Co-transfected LacZ served as control RNA. (c) Lane profile of the xrFrags shown in (b), indicating the position of xrFragA, B and C.



Supplementary Figure 3

NMD utilizes the whole spectrum of degradation pathways

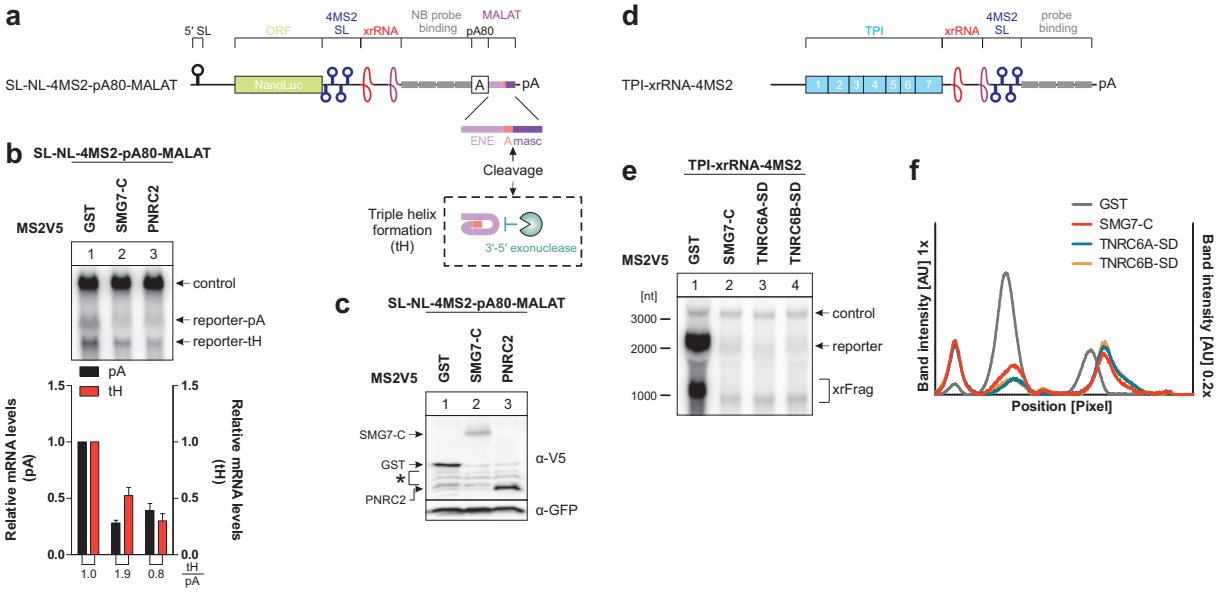
(a-c) Northern blots of RNA samples extracted from stable HeLa cell lines transfected with the indicated siRNAs and expressing the indicated reporter constructs. Endogenous 7SL served as control RNA. Mean values of reporter and xrFrag signal ± s.d. (n = 3) were quantified and for each knockdown condition the PTC values were normalized to the WT. The ratio of xrFrag to reporter mRNA levels is indicated below the graph. (d) Western blot analysis of knockdown efficiencies was done with the indicated antibodies, tubulin served as loading control.



Supplementary Figure 4

Degradation of IRES-containing dual xrRNA reporter via SMG6-catalyzed endocleavage

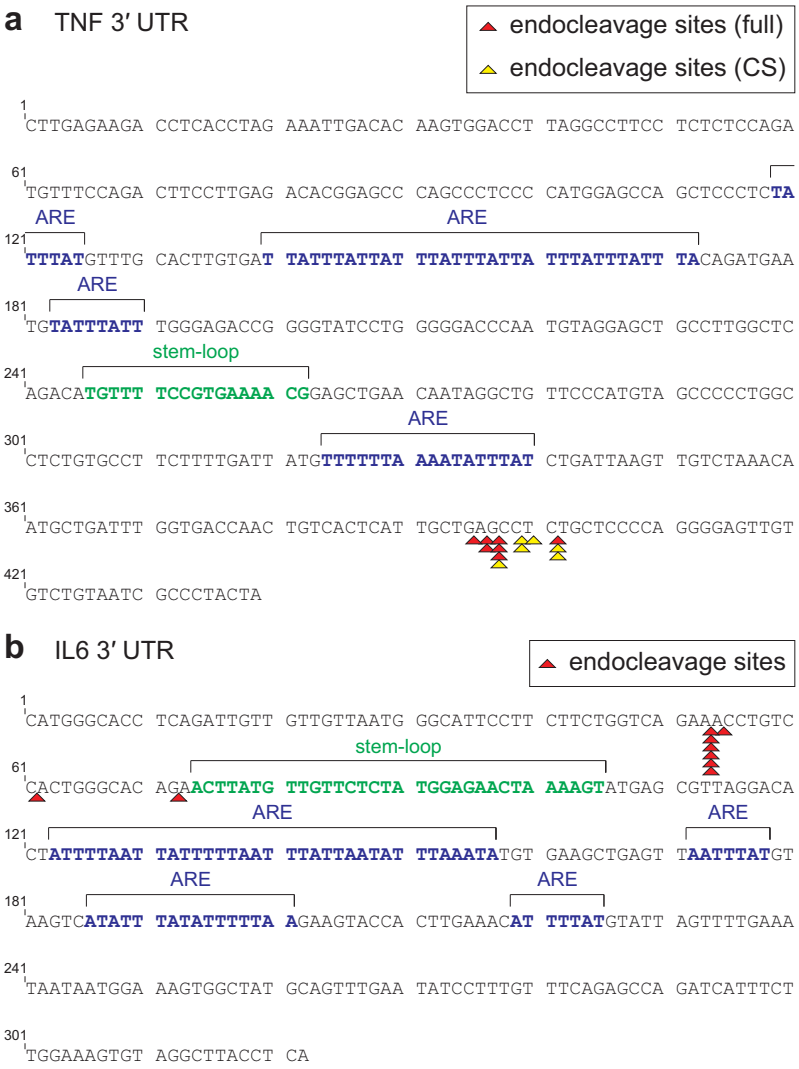
(a) Schematic representation of the dual xrRNA IRES reporter as in Fig. 1. (b,c) Northern blots of RNA samples extracted from HeLa cells transfected with the indicated siRNAs and reporter constructs. Co-transfected TPI (b) or globin WT (c) served as control mRNA. Mean values of reporter and xrFrag signals ± s.d. (n = 3) were quantified. The ratio of xrFrags to reporter mRNA levels is indicated below the graph.



Supplementary Figure 5

Tethered SMG7 induced deadenylation-dependent degradation of reporter mRNAs

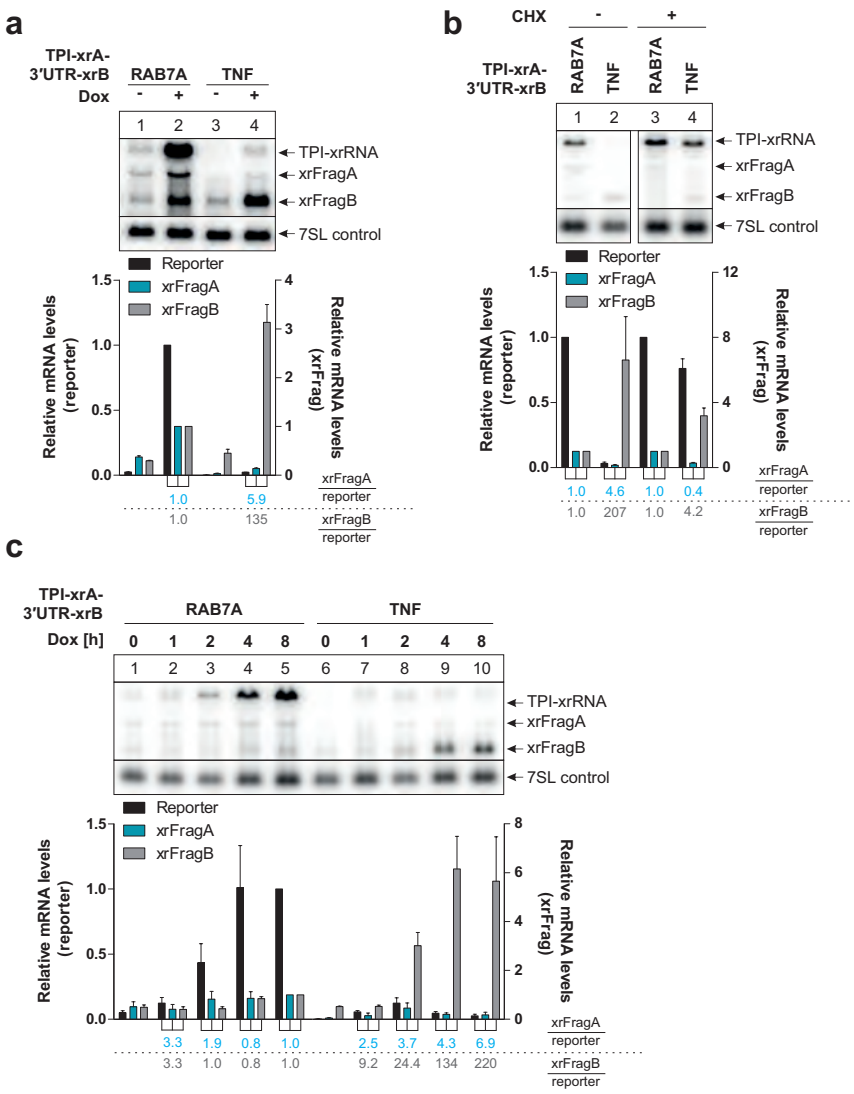
(a) Schematic representation of the non-translatable NanoLuc tethering reporter containing a 3′ MALAT sequence. To stimulate transport to the cytoplasm 80 template-encoded adenosines (pA80) were inserted upstream of the MALAT sequence. The processing by endonucleolytic cleavage of masc RNA and subsequent triple helix formation is indicated. (b) Northern blot of RNA samples extracted from HeLa cells transfected with the indicated tethering constructs and the SL-NL-4MS2-pA80-MALAT reporter. Mean values ± s.d. (n = 3) for polyadenylated reporter (pA) and triple helix reporter (tH) levels were quantified and normalized to tethered GST, which served as control. The ratio of tH:pA mRNA levels is indicated below the graph. (c) Western blot showing the expression levels of the MS2V5-tagged constructs with GFP serving as transfection control. Unspecific bands are indicated with asterisks. (d) Schematic representation of the TPI-xrRNA-4MS2 tethering reporter with inverted xrRNA and 4MS2 elements. (e) Northern blot of RNA samples extracted from HeLa cells transfected with the indicated tethering constructs and the TPI-xrRNA-4MS2 reporter, LacZ served as control. Approximate RNA size is indicated on the left. (f) Lane profiles of the northern blot in (e), with size-normalized LacZ control serving as reference point. Due to the differences in reporter and xrFrag intensities, the values for GST tethering were plotted on the left axis (1x intensity), for SMG7-C, TNRC6A-SD and TNRC6B-SD on the right axis (0.2x intensity).



Supplementary Figure 6

Annotated sequences of cytokine elements

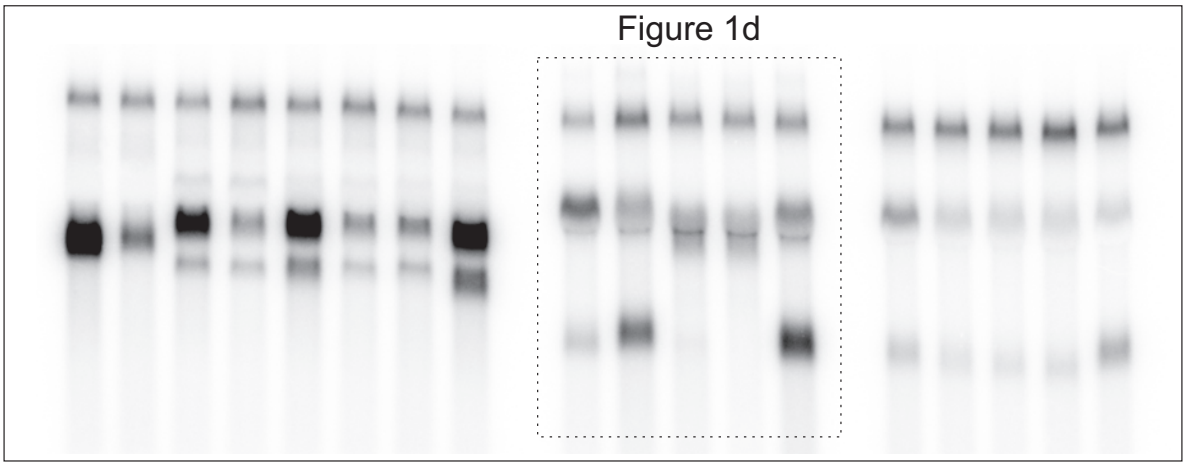
(a, b) The DNA sequence of the TNF-α and IL6 element used in reporter constructs are shown with annotations of sequence motifs. Single or overlapping AREs are annotated according to ARESite2[1]. Red and yellow triangles represent cleavage sites identified by sequencing of individual unique cloned 3′ fragments.

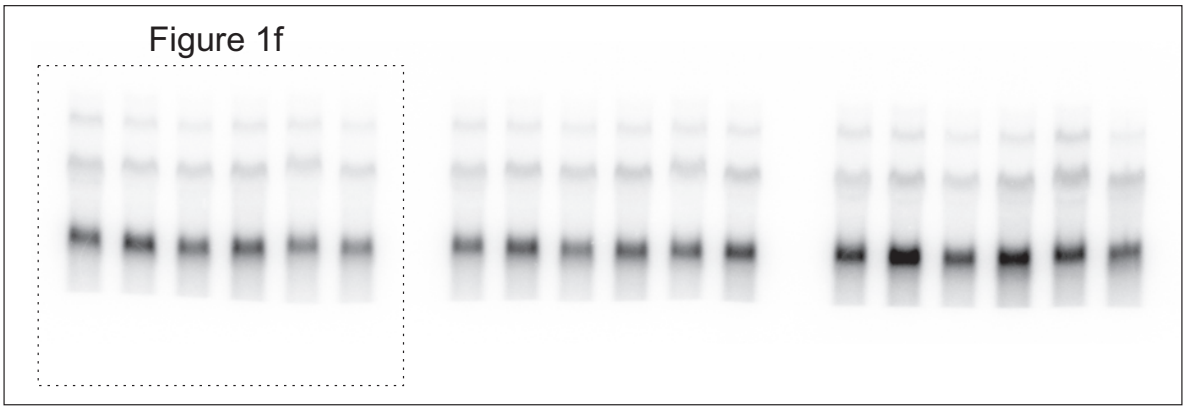


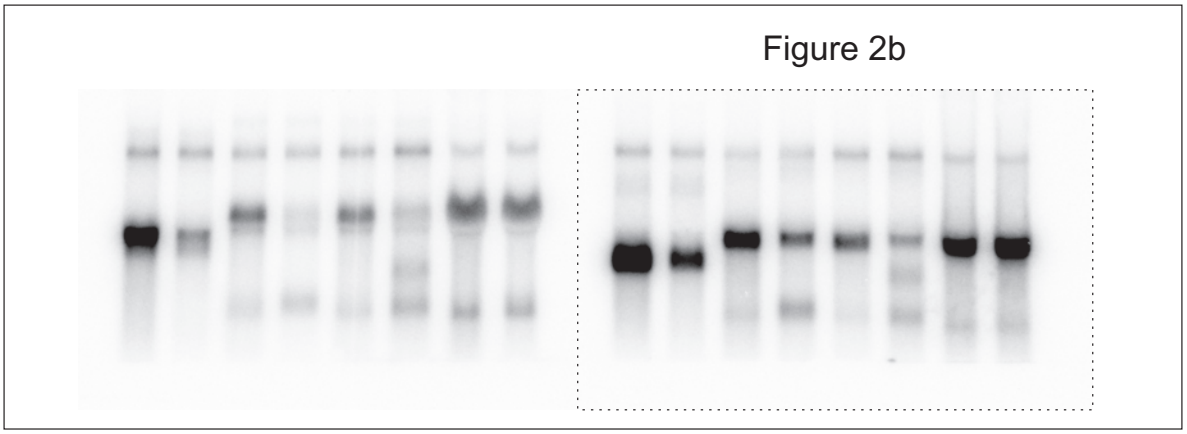
Supplementary Figure 7

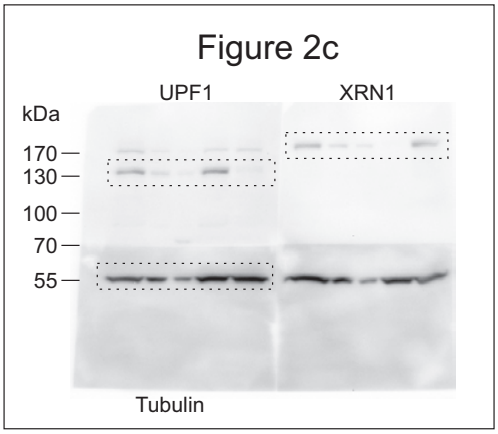
Detection of endocleavage within the TNF-α 3′ UTR framed by two xrRNA elements

(a-c) Northern blots of total RNA extracted from stable HeLa cell lines expressing the indicated dual xrRNA reporter mRNAs. Mean values of reporter and xrFrag signal ± s.d. (n = 3) were quantified and for each knockdown condition the PTC values were normalized to the WT. The ratio of xrFrag to reporter mRNA levels is indicated below the graph.



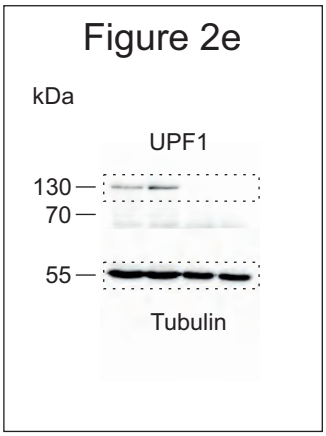


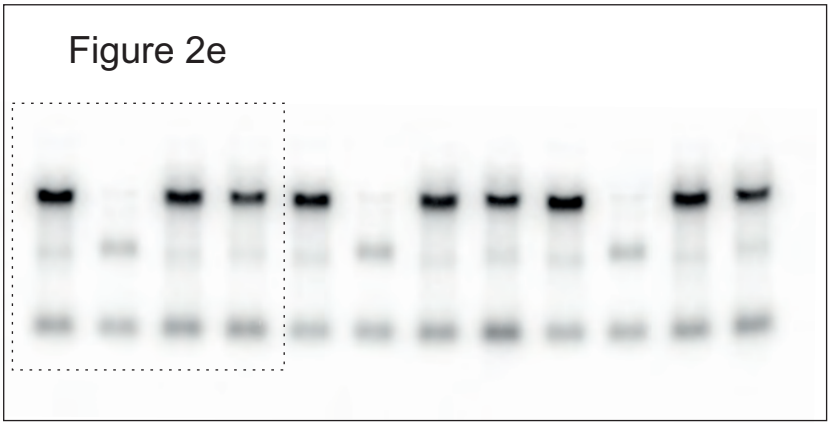


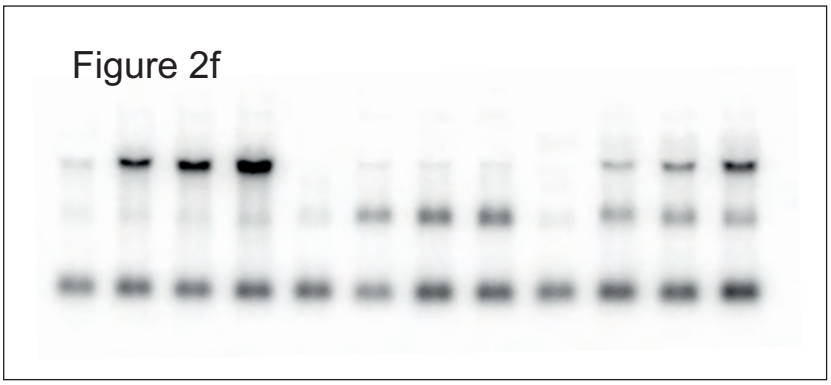


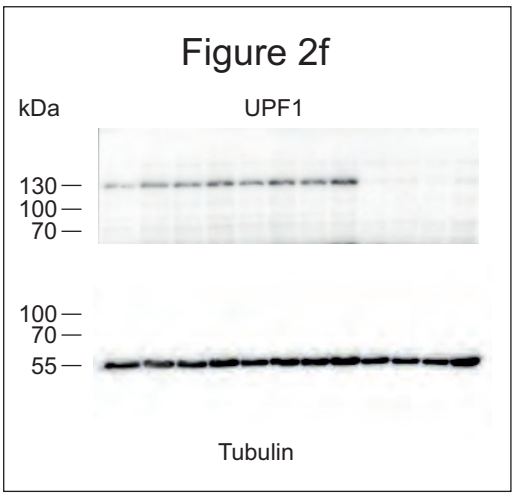
Supplementary Figure 8

Uncropped blots of Figures 1-8 as indicated.



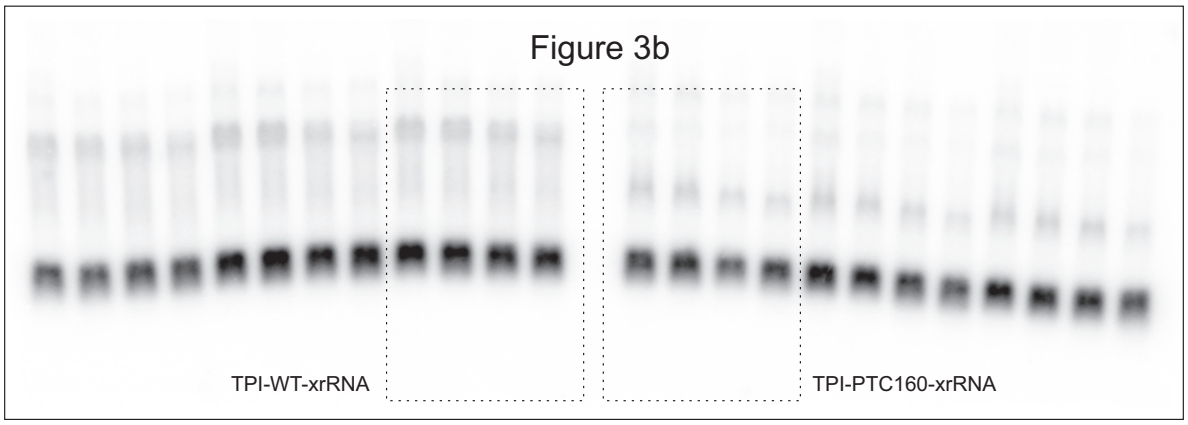


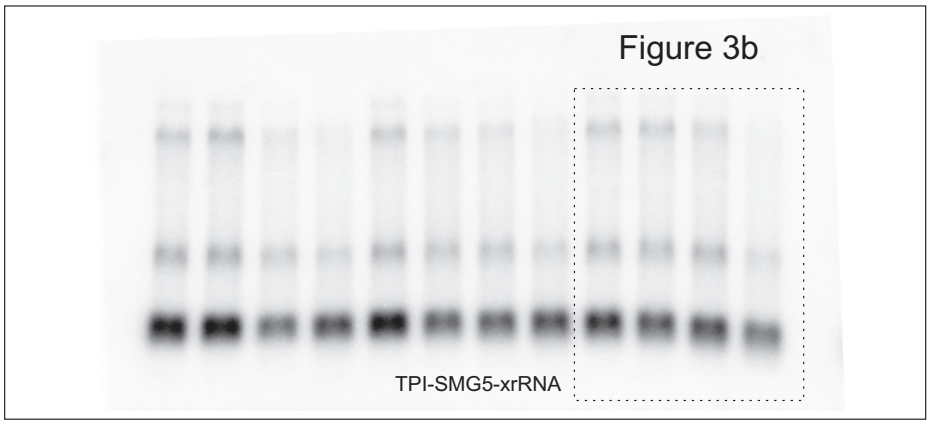


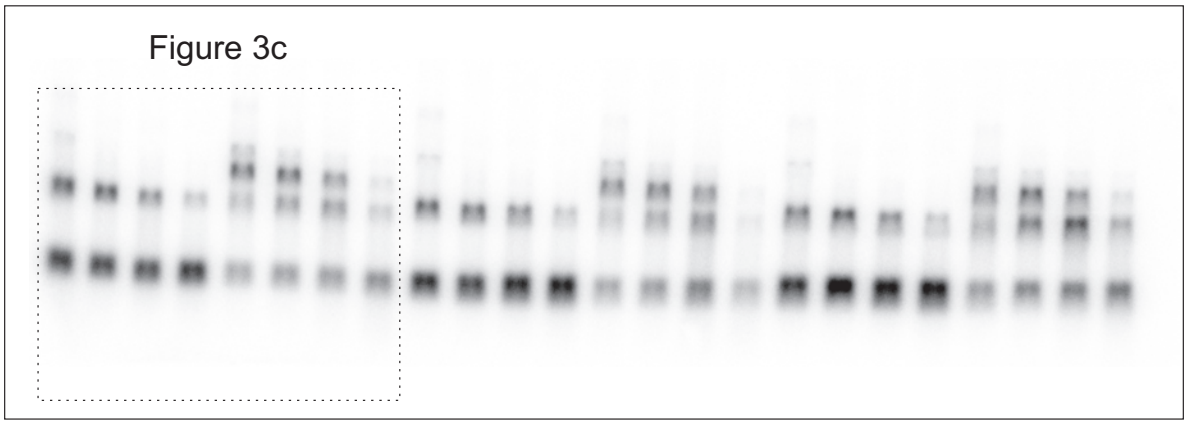


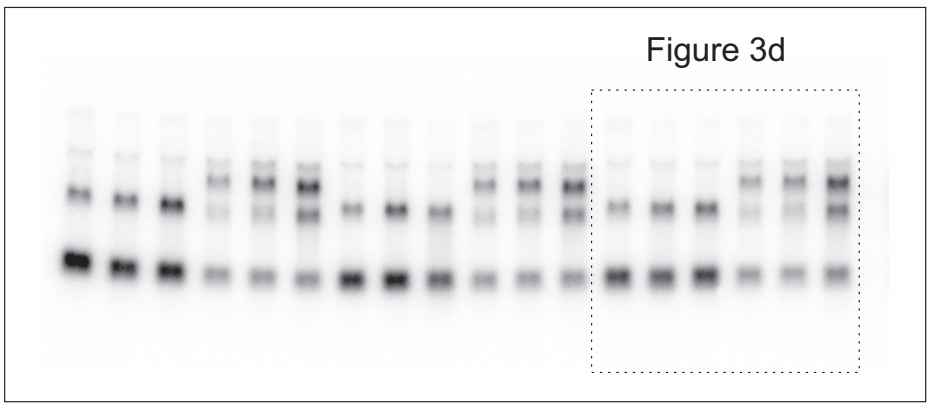
Supplementary Figure 8, continued

Uncropped blots of Figures 1-8 as indicated.



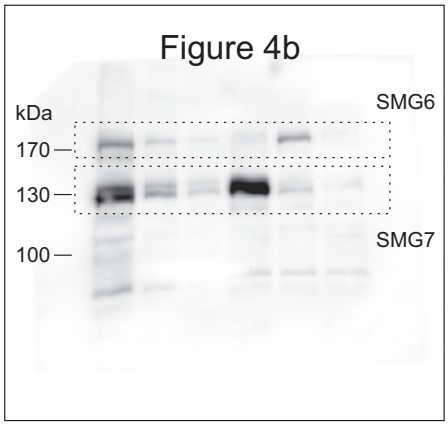


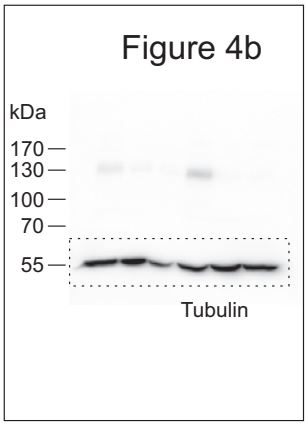


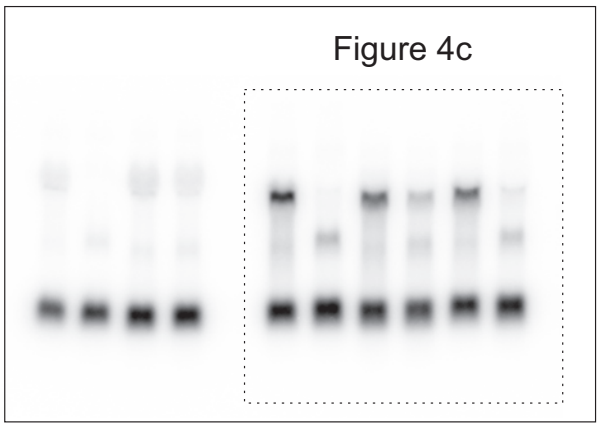


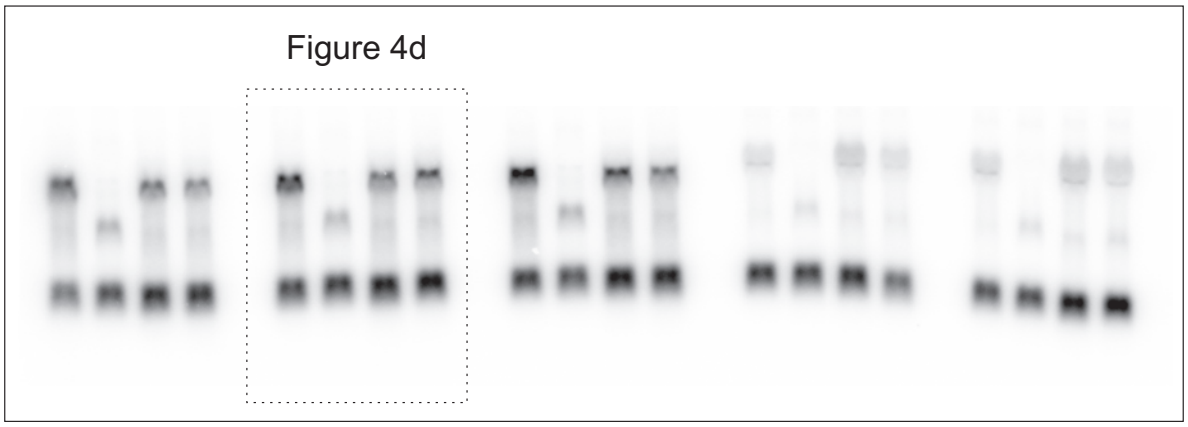
Supplementary Figure 8, continued

Uncropped blots of Figures 1-8 as indicated.



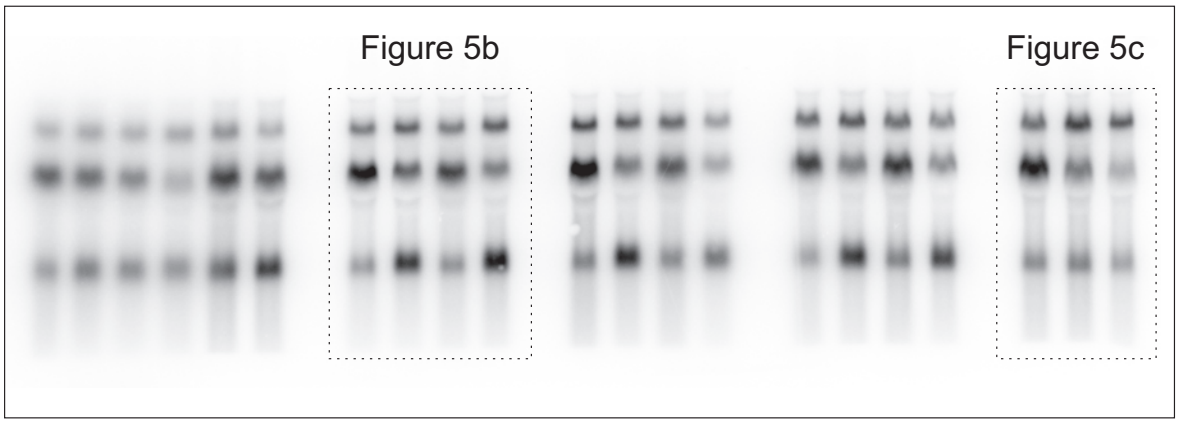


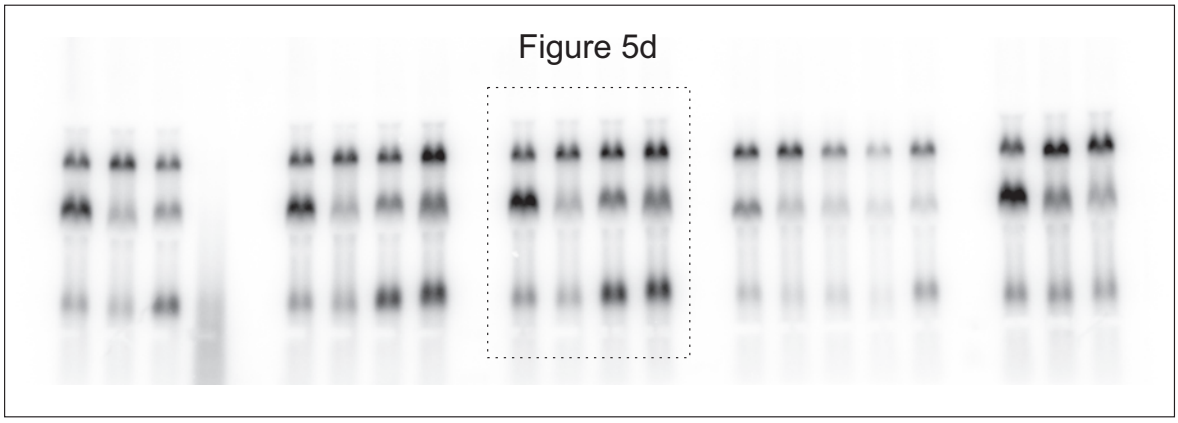


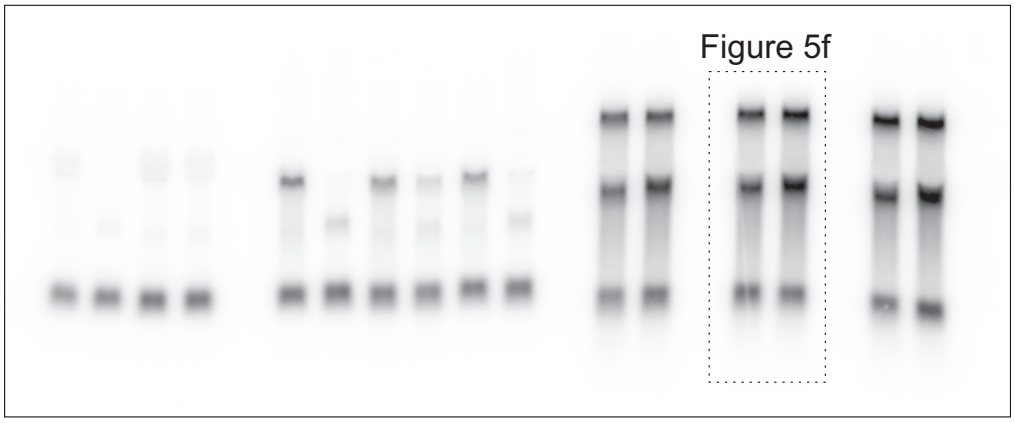


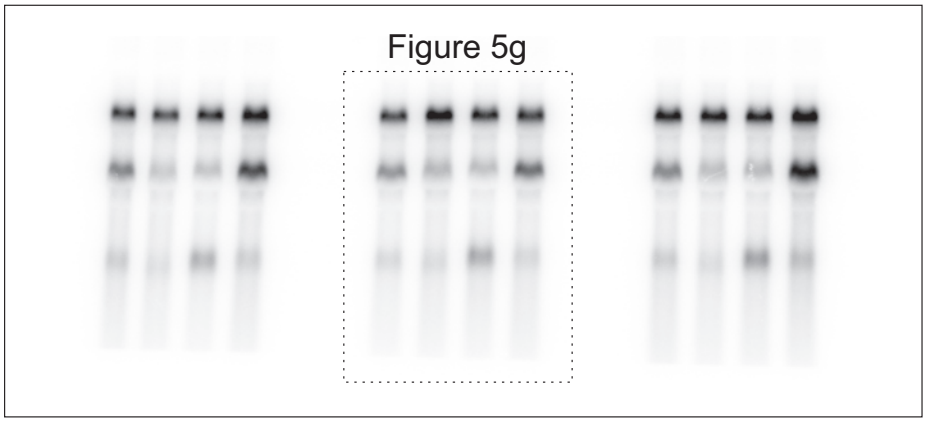
Supplementary Figure 8, continued

Uncropped blots of Figures 1-8 as indicated.



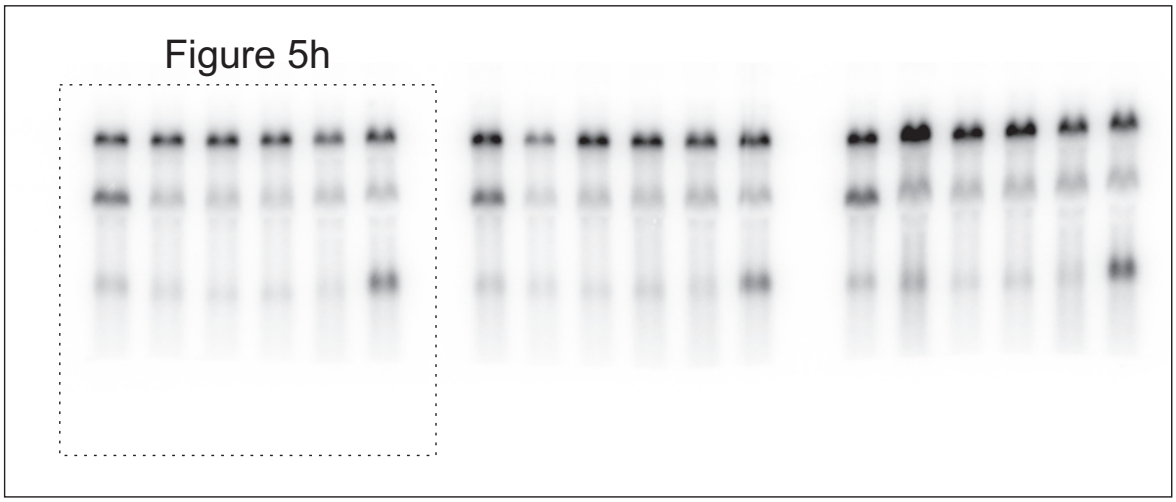


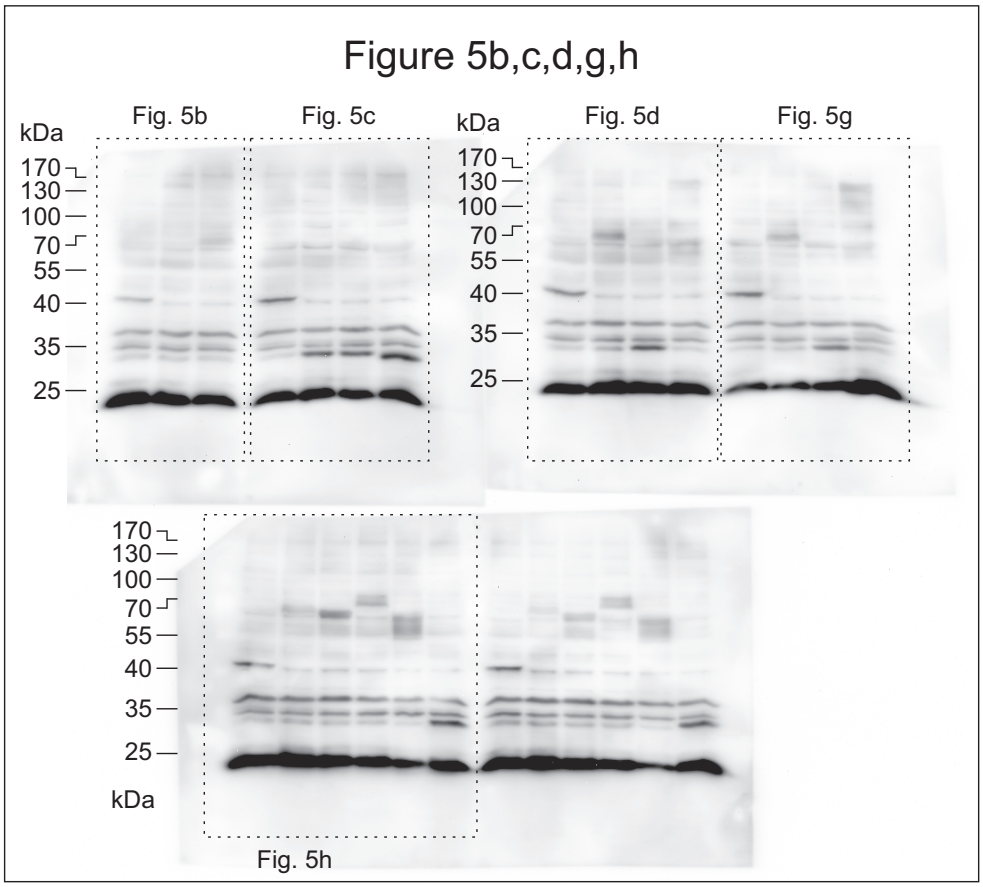




Supplementary Figure 8, continued

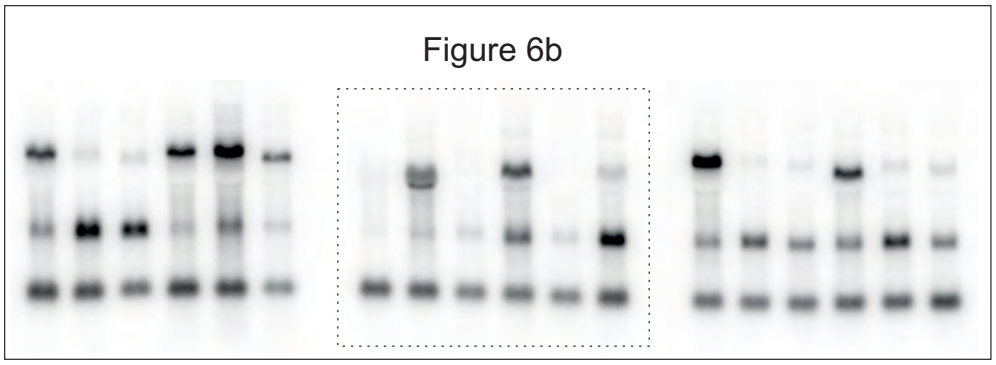
Uncropped blots of Figures 1-8 as indicated.

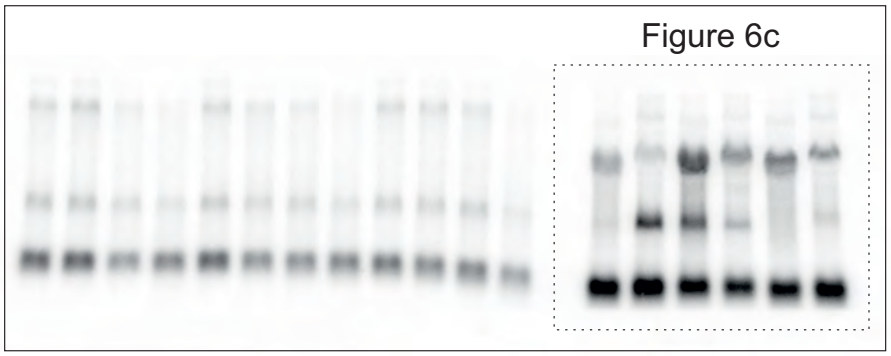


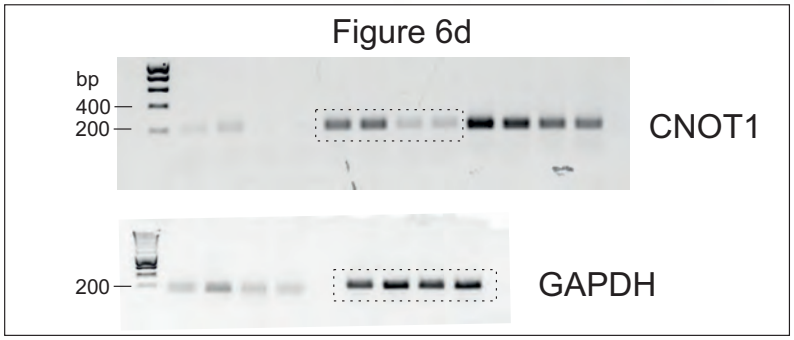


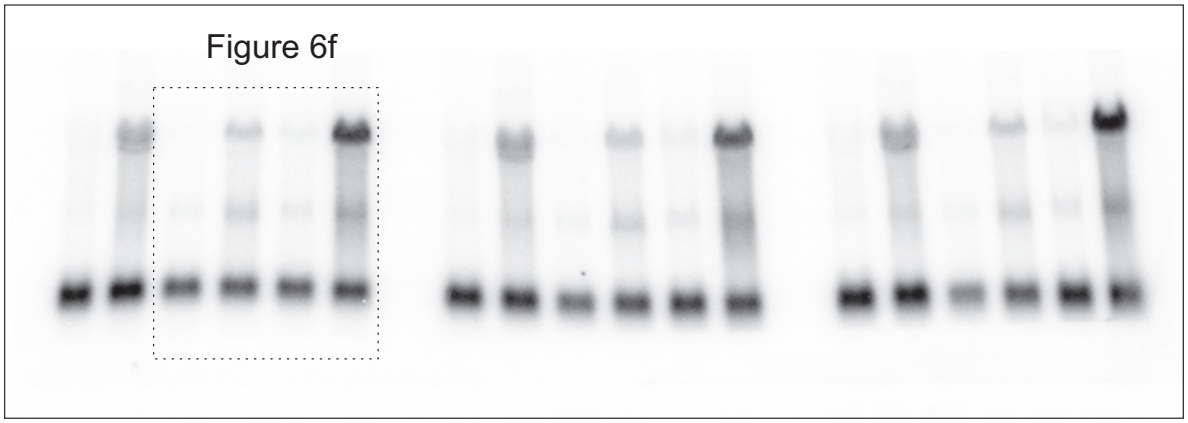
Supplementary Figure 8, continued

Uncropped blots of Figures 1-8 as indicated.



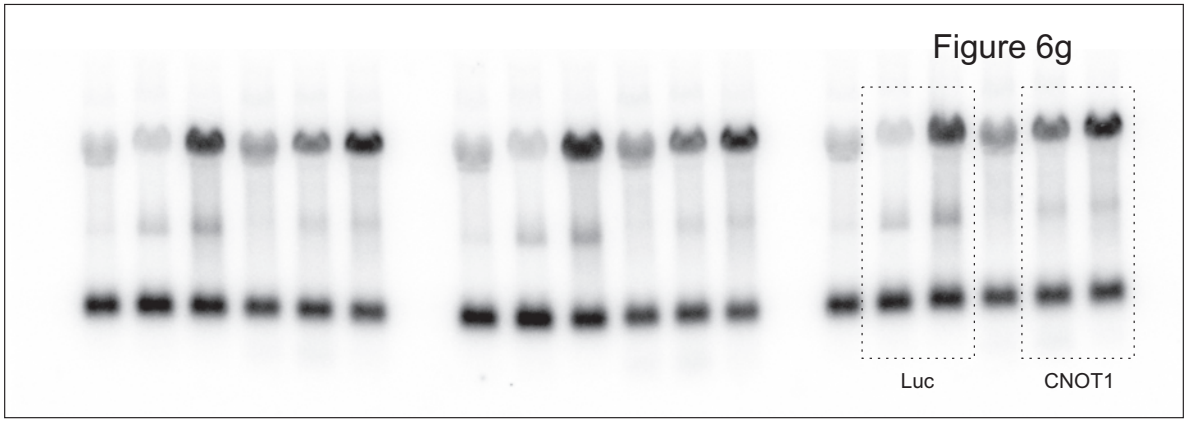


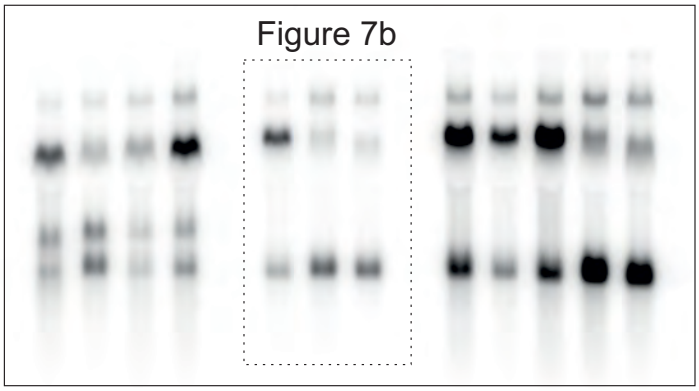


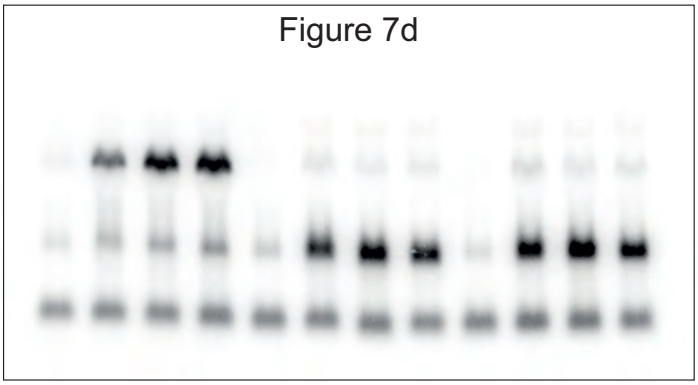


Supplementary Figure 8, continued

Uncropped blots of Figures 1-8 as indicated.

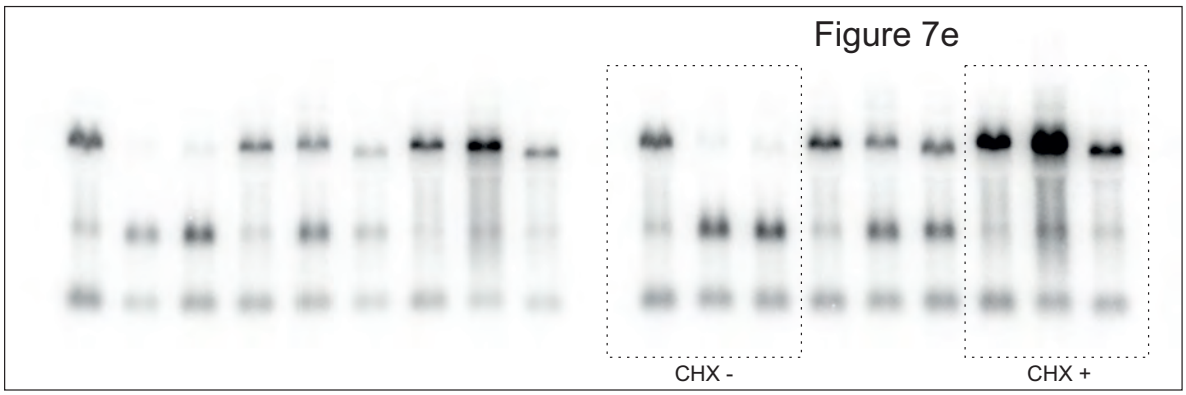


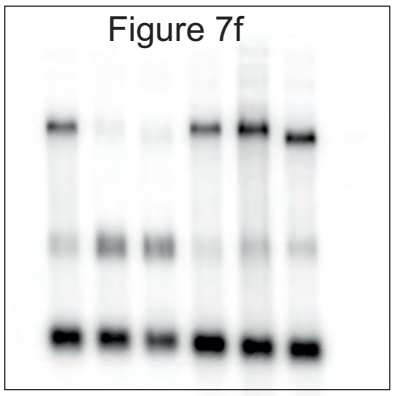


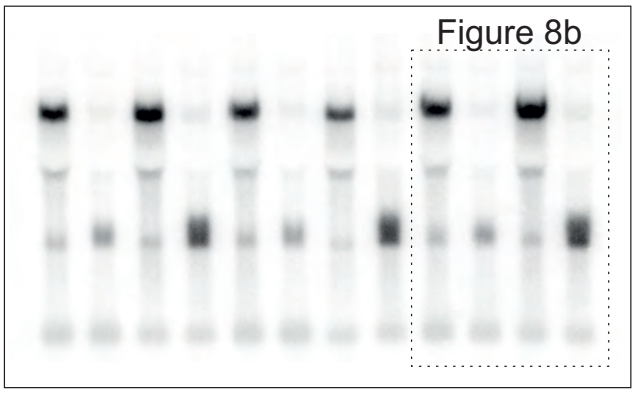


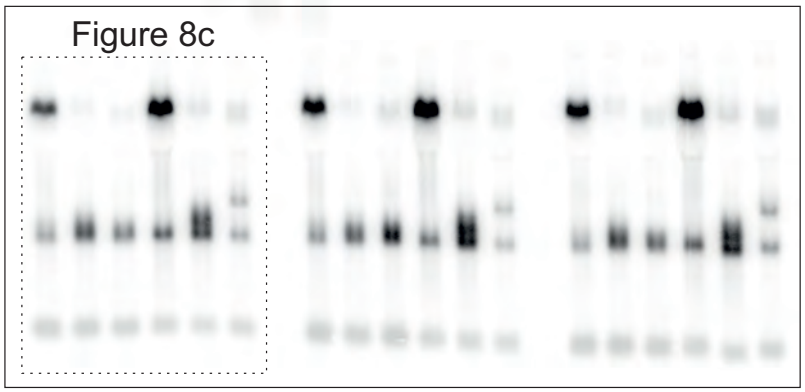
Supplementary Figure 8, continued

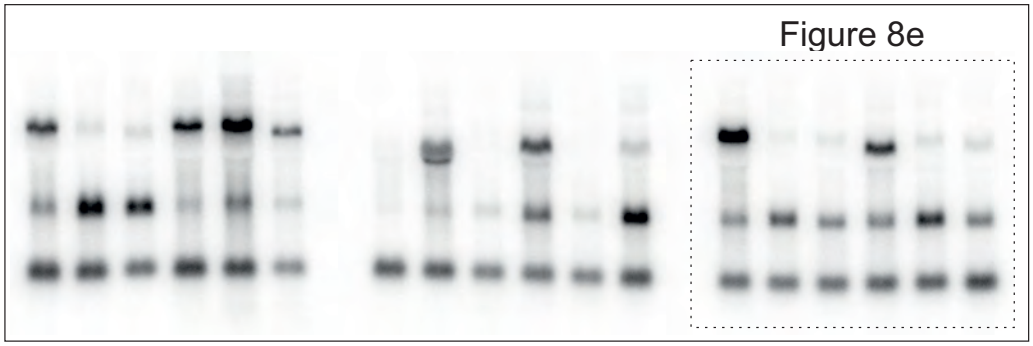
Uncropped blots of Figures 1-8 as indicated.





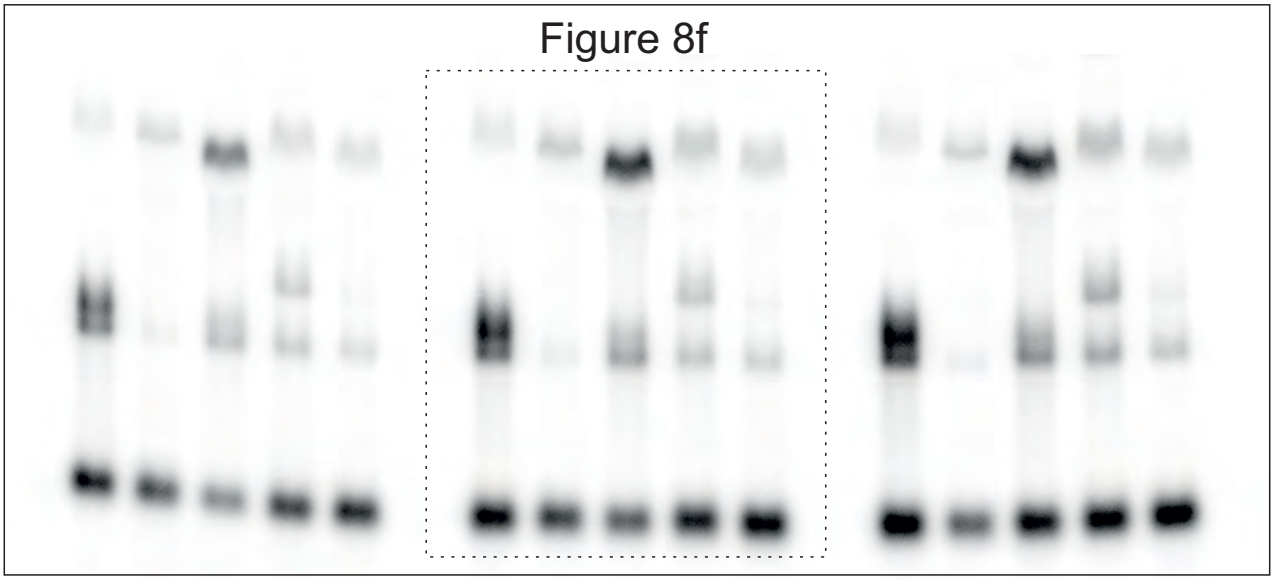






Supplementary Figure 8, continued

Uncropped blots of Figures 1-8 as indicated.



Supplementary Figure 8, continued

Uncropped blots of Figures 1-8 as indicated.

Supplementary References

1. Fallmann, J., Sedlyarov, V., Tanzer, A., Kovarik, P. & Hofacker, I.L. AREsite2: an enhanced database for the comprehensive investigation of AU/GU/U-rich elements. *Nucleic Acids Res* **44**, D90-5 (2016).

SCIENTIFIC REPORTS

OPEN

Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells

Received: 26 May 2017
Accepted: 15 August 2017
Published online: 06 September 2017

Jennifer V. Gerbracht, Volker Boehm & Niels H. Gehring

Messenger RNA (mRNA) turnover is a crucial and highly regulated step of gene expression in mammalian cells. This includes mRNA surveillance pathways such as nonsense-mediated mRNA decay (NMD), which assesses the fidelity of transcripts and eliminates mRNAs containing a premature translation termination codon (PTC). When studying mRNA degradation pathways, reporter mRNAs are commonly expressed in cultivated cells. Traditionally, the molecular mechanism of NMD has been characterized using pairs of reporter constructs that express the same mRNA with ("PTC-containing mRNA") or without ("wild-type mRNA") a PTC. Cell lines stably expressing an NMD reporter have been reported to yield very robust and highly reproducible results, but establishing the cell lines can be very time-consuming. Therefore, transient transfection of such reporter constructs is frequently used and allows analysis of many samples within a short period of time. However, the behavior of transiently and stably transfected NMD constructs has not been systematically compared so far. Here, we report that not all commonly used human cell lines degrade NMD targets following transient transfection. Furthermore, the degradation efficiency of NMD substrates can depend on the manner of transfection within the same cell line. This has substantial implications for the interpretation of NMD assays based on transient transfections.

The transfer of correct genetic information from the DNA to the translated protein is essential for the survival of a cell. Therefore, cellular quality control mechanisms exist, which detect and degrade aberrant transcripts[1]. Nonsense-mediated mRNA decay (NMD) is a translation-coupled quality control pathway in eukaryotic cells that targets mRNAs carrying a premature termination codon (PTC)[2]. Aberrant translation termination at the PTC causes stalling of the ribosome and the assembly of a surveillance complex, which recruits RNA degradation enzymes[3]. This prevents the production of truncated proteins with a potentially dominant-negative effect[4]. NMD is an important modulator of the phenotype of genetic disorders, as nonsense mutations represent 20% of disease-associated single base pair substitutions affecting gene coding regions[5]. Moreover, NMD acts as a general modulator of gene expression, since many transcripts that have arisen from alternative splicing or contain an upstream open reading frame or long 3′ UTR are endogenous NMD targets[6].

Numerous proteins and factors involved in NMD have been identified, but many questions regarding the exact mechanism by which NMD degrades mRNAs remain to be answered[7]. In the study of NMD, reporter mRNAs with and without a PTC are utilized to monitor the degradation of transcripts in standard human cell lines such as HeLa and HEK-293 cells. When introduced into the cell via transient transfection, these transcripts are expressed from a high number of extrachromosomal plasmids. Alternatively, isogenic cell lines stably expressing the reporter constructs can be established. When using a system such as Flp-In T-REx, the target gene is inserted at a single site in the cell line of choice. While overall expression levels of the reporter mRNAs are lower, previous studies suggest that more pronounced and robust results can be achieved using this approach[8]. However, generation of stable cell lines can be laborious and especially when a high number of reporter constructs are systematically compared, transient transfections yield faster results.

While studying NMD using many different cell lines, reporter constructs and transfection methods, we observed notable discrepancies. Since similar experimental designs are widely used in the field, we sought to systematically analyze the NMD efficiency in standard human cell lines. We found that not all commonly cultured

Institute for Genetics, Department of Biology, University of Cologne, 50674, Cologne, Germany. Correspondence and requests for materials should be addressed to N.H.G. (email: ngehring@uni-koeln.de)

human cells effectively degrade NMD targets when reporter mRNAs are expressed from extrachromosomal plasmid DNA. Additionally, for the widely used HEK-293 cells we observed differential NMD efficiency depending on whether stable or transient transfections have been performed.

Results and Discussion

Differential NMD efficiency in transiently transfected HeLa and HEK-293 cells. In order to determine the degradation efficiency of transfected (i.e. exogenous) NMD targets, we have developed an experimental system to compare different cell lines and transfection systems (Fig. 1a). First, we established inducible HeLa and HEK-293 Flp-In T-REx stable cell lines (referred to as HeLa FT and 293 FT, respectively) expressing triosephosphate isomerase (TPI) or globin reporter mRNA with or without a PTC. In both types of cells, the levels of PTC-containing full-length reporter mRNA were markedly reduced (Fig. 1b,c), indicating that these transcripts are efficiently degraded by NMD. An additional faster migrating band detected by northern blotting corresponds to a decay fragment (xrFrag) containing an xrRNA sequence, which is resistant to the 5′–3′ exonuclease XRN1[8]. The accumulation of xrFrags correlates with mRNA decay activity and serves as a readout for mRNA degradation pathways involving 5′–3′ decay[8]. When we transiently transfected the TPI and globin NMD reporter constructs in HeLa FT cells, the PTC-containing reporter mRNA was also strongly reduced (Fig. 1d). Strikingly, we did not observe differences between the WT and PTC constructs in transiently transfected 293 FT cells: the expression levels of both mRNAs were comparable and no xrFrags accumulated (Fig. 1e). This result was surprising since the reporter mRNA was expressed from the same pcDNA5/FRT/TO plasmid under control of the CMV promoter that was used for generating the stable cell lines. It has been reported before that NMD efficiency can vary between different strains of cells, for example for the HeLa cell model system[9] or epithelial cells of cystic fibrosis patients[10]. NMD efficiency has also been described to vary between different murine tissues[11]. In contrast, the observation made here is that the degradation of an NMD-target depends on the manner of transfection and not on the inherent characteristic of the cell.

NMD deficiency of transiently transfected HEK-293 cells is independent of transfection reagent and plasmid amount. Our initial observation was made using the calcium phosphate precipitation method for transient transfections. Although calcium phosphate-based transfections are commonly used and led to robust NMD readout in HeLa cells, we wanted to test whether the transfection method had an influence on the behavior of the reporter mRNA in the cell. To this end, the experiment was repeated using other commonly used reagents to transiently transfect cultured cells. However, no substantial differences between wild-type and PTC mRNAs in 293 FT cells were observed when the constructs were transfected using polyethylenimine (PEI) or Lipofectamine 2000 (Fig. 2a). Different cellular stress pathways have been previously implicated in the regulation of NMD efficiency[12]. To rule out that the observed inability to degrade an NMD reporter was due to cellular stress caused in response to transient transfections, a mock transfection of 293 FT cells stably expressing globin PTC39 mRNA was performed (Fig. 2b). When the stable cell lines were transiently co-transfected with salmon sperm DNA, the mRNA expressed from the stably integrated PTC39 construct was still efficiently degraded (Fig. 2b, lane 3). The same observation was made when the cells were co-transfected with the lacZ expression plasmid (Fig. 2b, lane 4) as well as with a GFP-expressing plasmid (Fig. 2b, lane 5, GFP expression was confirmed by microscopy), which were both included in our transient transfections as controls.

For functional NMD, the transcript must be exported to the cytoplasm and be actively translated. We confirmed that the reporter mRNA is exported to the cytoplasm via subcellular fractionation (Fig. 2c). Furthermore, the translatability of these transcripts was assessed by expressing GFP from transiently transfected plasmids in 293 FT cells and confirming GFP levels by microscopy.

In stably transfected FT cells, the gene of interest is integrated at a single site in the genome in every cell and the total cell population is isogenic. In transiently transfected cells however, the amount of plasmid DNA and expressed mRNA varies between cells. To exclude that the apparent inhibition of NMD results from overloading individual cells using the transient system, we transfected 293 FT cells with lower amounts of plasmid DNA. We estimated that 0.1–0.3 µg of transiently transfected plasmid DNA yields expression levels that are comparable to the amount expressed in stable cell lines after 24 h of induction (Fig. 3a). When transfecting 293 FT cells using these reduced concentrations, still no degradation of the PTC-containing mRNA could be detected down to 0.1 µg of transfected plasmid DNA (Fig. 3b). Although the average mRNA levels in stable cell lines and cells transfected with 0.1 µg plasmid DNA were comparable, individual transiently transfected cells might express more than the average stably transfected cell. Therefore, we transfected lower amounts of DNA (0.01 and 0.03 µg) and quantified reporter mRNA expression by qPCR, since expression levels were too low to be detectable by northern blotting (Fig. 3c). Even when using these minimal amounts of transfected reporter DNA, no difference could be detected between wild-type and PTC-containing reporter mRNA.

Collectively, these results show that the inability of transiently transfected 293 FT cells to degrade NMD reporter mRNAs is independent of transfection reagent and the amount of transfected plasmid DNA. This is in line with the fact that HeLa cells, which have been transiently transfected with a high amount of plasmid DNA, degrade PTC-containing reporters very efficiently (Fig. 1d).

Not all degradation pathways are impaired in transiently transfected HEK-293 cells. To investigate whether 293 FT cells exhibit a general defect to efficiently degrade unstable mRNAs expressed from transiently transfected plasmids, we studied reporter mRNAs containing the 3′ UTRs from the cytokines TNF-α and Interleukin-6 (IL6) in HeLa and 293 FT cells (Fig. 4a). These 3′ UTRs contain several degradation-promoting sequences such as AU-rich elements and a decay-inducing stem-loop[13]. Interestingly, both reporter mRNAs were

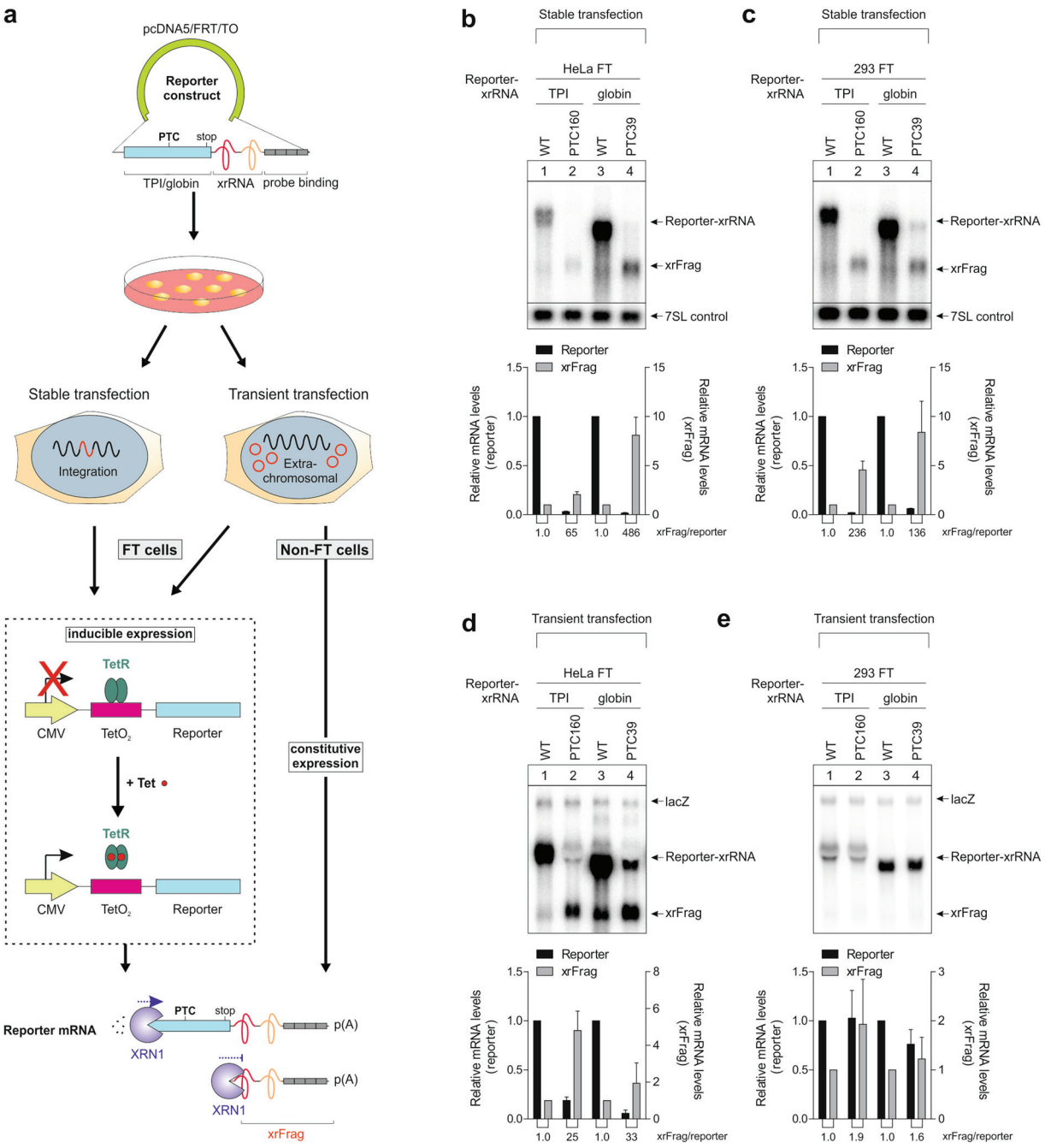


Figure 1. Transiently transfected 293 FT cells show a strongly reduced NMD efficiency in comparison to stably transfected cells. (**a**) Schematic representation of the experimental setup. The reporter constructs are generated by cloning the gene of interest into the pcDNA5/FRT/TO vector. The blue bar represents the coding sequence of triosephosphate isomerase (TPI) or β-globin with or without a premature termination codon (PTC) and with a normal stop codon (stop). Downstream, a viral XRN1-resistant sequence (xrRNA) is present that allows the detection of decay intermediates resulting from 5′–3′ exonucleolytic decay. The grey boxes correspond to four repeats of a northern blot probe binding site. Cells are transfected with these constructs using either a stable or transient transfection system. Flp-In T-REx (FT) cells constitutively express a Tet repressor (TetR) that blocks transcription of the reporter by binding to Tet operators ($TetO_2$) present downstream of the cytomegalovirus (CMV) promoter in the reporter construct. Therefore, FT cells are induced with tetracycline(Tet)/doxycycline. In non-FT cells, the reporter is always expressed. Unstable reporter mRNAs are degraded with participation of the cellular 5′–3′ exonuclease XRN1. Full-length reporter levels and decay fragment (xrFrag) caused by stalling of XRN1 at the xrRNA are detected via northern blotting. p(A): poly(A) tail. (**b**,**c**) Expression of reporter mRNAs in HeLa FT (**b**) and 293 FT (**c**) stable cell lines was induced with doxycycline for 24 h and the reporter levels detected by northern blotting. The lower band (xrFrag) corresponds to a 5′–3′ exonucleolytic decay intermediate. Endogenous 7SL RNA levels were detected and quantified on the same blot and are shown as control. The mean values ± SD (n = 3) for relative transcript levels were quantified and the PTC values normalized to the wild-type control. The xrRNA/reporter ratio is indicated below the graph. (**d**,**e**) HeLa FT

(**d**) and 293 FT (**e**) cells were transiently transfected with 3 µg plasmid DNA per 6-well and transcription was induced with doxycycline for 24 h. Reporter mRNA levels were analyzed by northern blotting. LacZ was co-expressed and serves as a transfection control. The mean values ± SD (n = 3) were quantified and the PTC values normalized to the wild-type control. The xrRNA/reporter ratio is shown below the graph.

degraded in 293 FT cells, although the reduction of full length mRNA as well as accumulation of xrFrags was less pronounced than in HeLa FT cells (Fig. 4b,c).

Variable degradation of transiently transfected NMD reporters in other human cell lines. In order to establish whether the phenomenon of impaired NMD efficiency of transiently transfected cells is limited to 293 FT cells, more human cell lines were analyzed (Fig. 5). First, other strains of HeLa and HEK-293 cells were examined, which recapitulated the results obtained from their Flp-In T-REx counterparts (Fig. 5a,b). Next, further commonly used cancer cell lines were investigated. In U2OS cells that have been transiently transfected with globin reporter mRNA with or without a PTC, the full-length reporter levels were reduced, whereas the xrFrag accumulated, suggesting efficient NMD (Fig. 5c). In contrast, in MCF-7 cells transfected with the same construct, the full-length reporter mRNA levels were not reduced (Fig. 5d), analogous to the observation made in HEK-293 cells. These findings demonstrate that reduced NMD of transiently expressed reporter mRNAs is not limited to HEK-293 cells and that caution should be taken when investigating NMD in human cell lines using a transient transfection system.

Conclusion

To the best of our knowledge, this is the first comparison of stably vs. transiently transfected NMD reporters in cultured human cells. Of the six different cells lines tested in this study, three (293 FT, 293 MSR and MCF-7) did not show efficient degradation of NMD-reporter mRNAs in transient transfection experiments. Importantly, HEK-293 and MCF-7 cells are widely used and easy to handle human cell lines and may therefore be chosen for transient transfections by a substantial number of laboratories. Considering that we have analyzed only a small number of cell lines until now, it may be premature to conclude that NMD substrates are generally less efficiently degraded when expressed from transiently transfected reporter constructs. Nonetheless, our results raise the concern that the robustness of NMD assays using transfected reporters may be influenced by a previously unknown interference of transient transfections with the NMD competence of certain cell lines. Currently, the reason for the impaired NMD of transiently transfected reporters in HEK-293 cells and MCF-7 cells is unclear. We have confirmed that mRNAs expressed from extrachromosomal plasmids are transported to the cytoplasm and efficiently translated, two known requirements for the activation of NMD. It is conceivable that transiently expressed mRNAs are unable to recruit NMD-promoting messenger ribonucleoprotein (mRNP) components. However, the identity of such mRNP components remain to be determined in future experiments. Furthermore, we cannot exclude that transiently expressed mRNAs are in general more resistant to mRNA degradation in HEK-293 cells. Although the decay of TNF-α and Interleukin-6 3′ UTR-containing reporter mRNAs is more pronounced than that our NMD reporter mRNAs, it not does not match the turnover efficiency of reporter mRNAs expressed from stably integrated constructs. Taken together, mRNAs/mRNPs expressed by transient transfections may exhibit a differential molecular composition or contain certain modifications that render them more resistant to mRNA turnover than mRNAs transcribed from chromosomal insertions. Therefore, we strongly suggest to assess NMD efficiencies using independent methods. Such initial tests will be particularly important when previously uncharacterized cell lines are used or when NMD efficiencies of different cells are compared.

Materials and Methods

Plasmid constructs and cell culture. The plasmid constructs β-globin-WT and -PTC39; TPI-WT and -PTC160; TPI-RAB7A, -TNF and -IL6 containing an XRN1-resistant structure from the Murray Valley encephalitis virus were generated by cloning the respective DNA fragments[8] into the pcDNA5/FRT/TO vector (Thermo Fisher Scientific). The mVenus and lacZ control plasmid were constructed by inserting the respective DNA fragments into the pCI-neo vector (Promega) as described previously[14].

HEK-293 Flp-In T-REx (293 FT; Thermo Fisher Scientific), HeLa Flp-In T-REx (HeLa FT; established by Elena Dobrikova and Matthias Gromeier, Duke University Medical Center), HeLa Tet-Off (Clontech), GripTite 293 MSR (Thermo Fisher Scientific), U2OS and MCF-7 cells were cultured in Dulbecco's Modified Eagle's Medium (Gibco, Life Technologies) supplemented with Penicillin-Streptomycin and 9% FBS. The cells were maintained at 37 °C, 5% CO_2 and 90% humidity.

Establishment of stable cell lines. For stable transfections, HeLa FT or 293 FT cells were seeded at a density of 2.8×10^5 cells/well in 6-well plates 24 h before transfection. In each well, 2.5 µg of the pcDNA5/FRT/TO vector with the gene of interest and 0.5 µg of the pOG44 vector expressing the Flp recombinase were transfected with the calcium phosphate method. 48 h after transfection, the cells were transferred to 10 cm culture dishes and selected with a hygromycin concentration of 100 µg/ml (293 FT cells) or 150 µg/ml (HeLa FT cells). Expression of stable cell lines was induced with 1 µg/ml doxycycline for 24 h.

Transient plasmid transfections. 2.8×10^5 cells were seeded in 6-well plates 24 h before transfection. If not indicated otherwise, the cells were transfected using the calcium phosphate method. The indicated amount of reporter plasmid was co-transfected with 0.5 µg of the mVenus expression plasmid and 3 µg of the lacZ control plasmid. Polyethylenimine (PEI) transfections were performed according to the jetPEI protocol (Polyplus transfection). Transfections using jetPRIME (Polyplus transfection) and Lipofectamine 2000 (Thermo Fisher

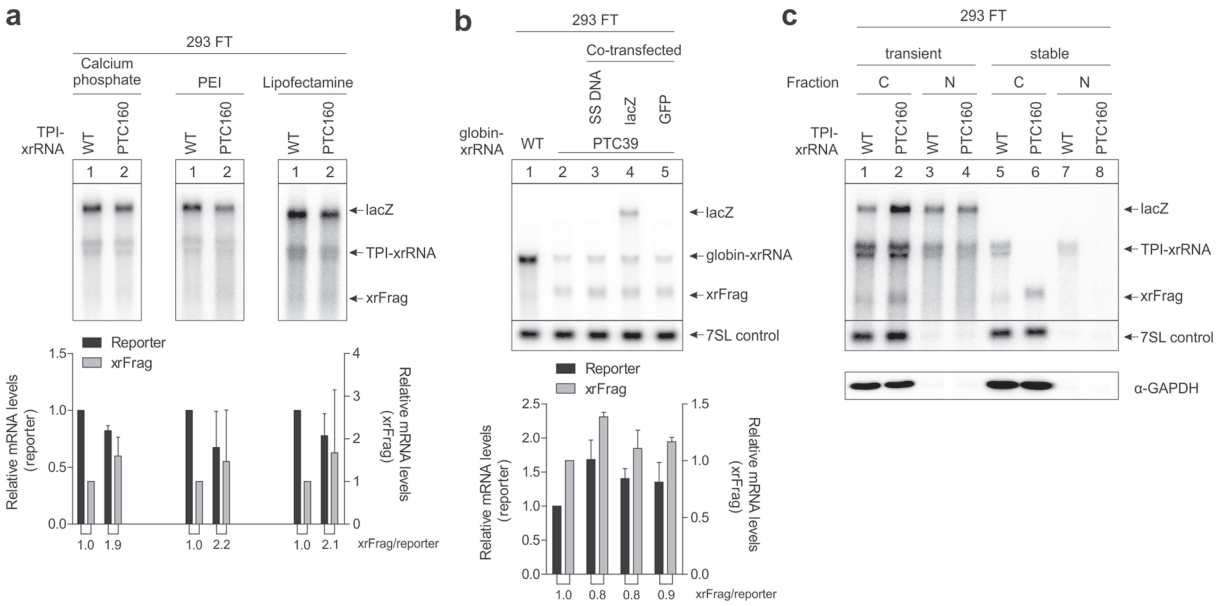


Figure 2. The reduced NMD efficiency in transiently transfected 293 FT cells is not due to transfection conditions. (**a**) Northern blots of the indicated reporter mRNAs. 293 FT cells were transiently transfected with 0.3 µg of plasmid DNA per 6-well using different transfection reagents. LacZ was co-transfected as a control mRNA. The mean reporter values ± SD (n = 3) were quantified and the PTC-reporter levels normalized to the wild-type control. The xrRNA/reporter ratio is indicated below the graph. (**b**) RNA samples were extracted from stably transfected 293 FT cells and analyzed by northern blotting. The samples shown in Lanes 3–5 have additionally been transiently transfected with salmon sperm DNA (SS DNA), lacZ and a GFP expression plasmid, respectively. Endogenous 7SL RNA levels were detected and quantified on the same blot and are shown as control. The mean values ± SD (n = 3) were quantified and the values of the co-transfected samples normalized to the untransfected PTC control. The xrRNA/reporter ratio is indicated below the graph. (**c**) Northern blot of RNA extracted from the indicated fractions of 293 FT cells transfected with 0.3 µg of plasmid DNA per 6-well. Co-transfected lacZ is shown as a control. Endogenous 7SL RNA levels were detected on the same blot as control. C: cytoplasmic fraction, N: nuclear fraction. Successful fractionation was confirmed by the cytoplasmic marker GAPDH and the localization of the 7SL control RNA.

Scientific) were performed according to the manufacturer's instructions. Due to the presence of Tet repressor binding sites in the pcDNA5/FRT/TO vector and the constitutive expression of the Tet repressor in the FT cell lines, expression from transiently transfected pcDNA5/FRT/TO reporter plasmids in these cells was induced by supplementing the medium with 1 µg/ml doxycycline for 24 h.

Subcellular fractionation. The cells were harvested in polysome buffer (10 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris-HCl pH 7.4, 1% Triton X-100, 1% Na-deoxycholate and 1 mM DTT). The cytoplasmic and nuclear fractions were separated by centrifugation.

RNA extraction and northern blotting. The cells were harvested in peqGOLD TriFast reagent (VWR) and total RNA extraction was performed as recommended by the manufacturer's protocol. 2.5 µg of total RNA were resolved on a 1% agarose/0.4 M formaldehyde gel using the tricine/triethanolamine buffer system as described[15] followed by transfer on a nylon membrane (Roth) in 10x SSC. The blots were incubated overnight at 65 °C in Church buffer containing [α-^{32}P]-GTP body-labeled RNA probes for detection of the reporter and lacZ control RNA. Endogenous 7SL RNA was detected by a 5′-^{32}P-labeled oligonucleotide (5′-TGCTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTT-3′). The blots were visualized and quantified using the Typhoon FLA 7000 (GE Healthcare) and ImageQuant TL 1D software.

Western blot analysis. Proteins were extracted using peqGOLD TriFast reagent (VWR), separated by SDS-PAGE gel electrophoresis and transferred to nitrocellulose membranes. The following antibodies were used: anti-GAPDH (Santa Cruz Biotechnology), peroxidase-coupled secondary anti-mouse (Jackson ImmunoResearch). Detection was performed with Western Lightning Plus-ECL (PerkinElmer) and myECL Imager (Thermo Fisher Scientific).

Quantitative real-time PCR analysis. 1 µg of total RNA was reverse-transcribed in a 20 µl reaction volume with an oligo dT primer (5′-TTTTTTTTTTTTTTTTTTTTVNN-3′, 10 µM final concentration) using the ProtoScript II Reverse Transcriptase (NEB) according to the manufacturer's instructions. PCRs were performed using one-tenth of the reverse transcription reaction with GoTaq qPCR Master Mix (Promega) and the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). The reactions were performed in triplicates and the average C_t

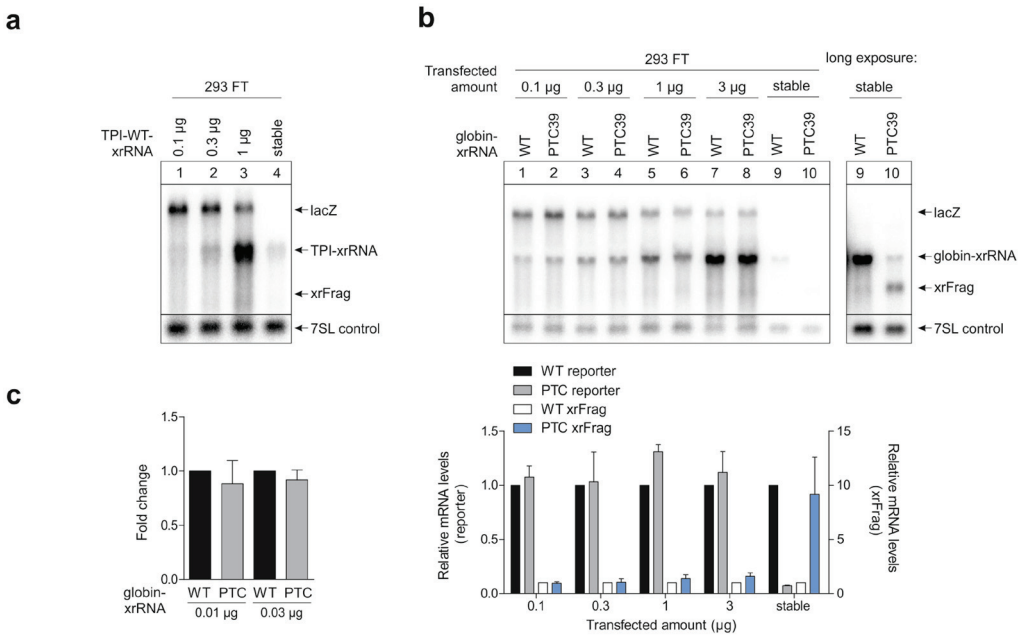


Figure 3. NMD efficiency in transiently transfected 293 FT cells does not depend on the amount of transfected plasmid DNA. (**a**) Lanes 1–3: 293 FT cells were transfected with the indicated amount of plasmid DNA per 6-well and the mRNA was analyzed by northern blotting. LacZ was co-transfected as a control. Lane 4: 293 FT cells stably transfected with the indicated construct were induced for 24 h. Endogenous 7SL RNA levels were detected on the same blot and are shown as control. (**b**) Lanes 1–8: 293 FT cells were transfected with the indicated amount of plasmid DNA per 6-well and the mRNA was analyzed by northern blotting. LacZ was co-transfected as a control. Lanes 9–10: 293 FT cells stably transfected with the indicated construct were induced for 24 h. Endogenous 7SL RNA levels were detected and quantified on the same blot and are shown as control. The mean values ± SD (n = 3) of the signals were quantified and the PTC-values normalized to the corresponding wild-type control. (**c**) 293 FT cells were transiently transfected with the indicated amount of plasmid DNA per 6-well and expression levels were measured by qPCR. The reporter mRNA levels were normalized to the co-transfected GFP control. The bars represent mean fold changes of PTC-containing reporter vs. wild-type reporter levels ± SD (n = 3).

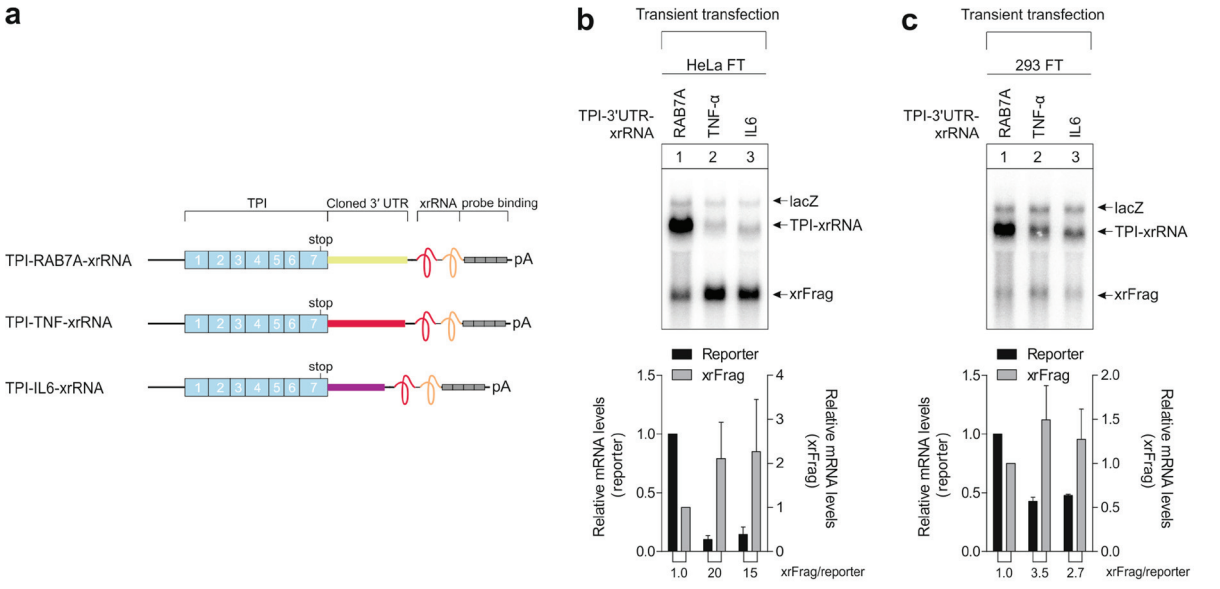


Figure 4. Degradation of unstable reporter mRNAs in 293 FT cells. (**a**) Schematic representation of the reporter constructs as in Fig. 1a. Downstream of the TPI gene, the 3′ UTRs of RAB7A (control), TNF-α and IL6 were cloned. (**b**,**c**) HeLa FT (**b**) and 293 FT (**c**) cells were transfected with 3 µg plasmid DNA of the indicated reporter per 6-well. The reporter levels were analyzed by northern blotting. LacZ was co-transfected as a transfection control. The mean reporter values ± SD (n = 3) were quantified and the cytokine reporter mRNAs normalized to the RAB7A control. The xrFrag/reporter ratio was calculated and is indicated below the graph.

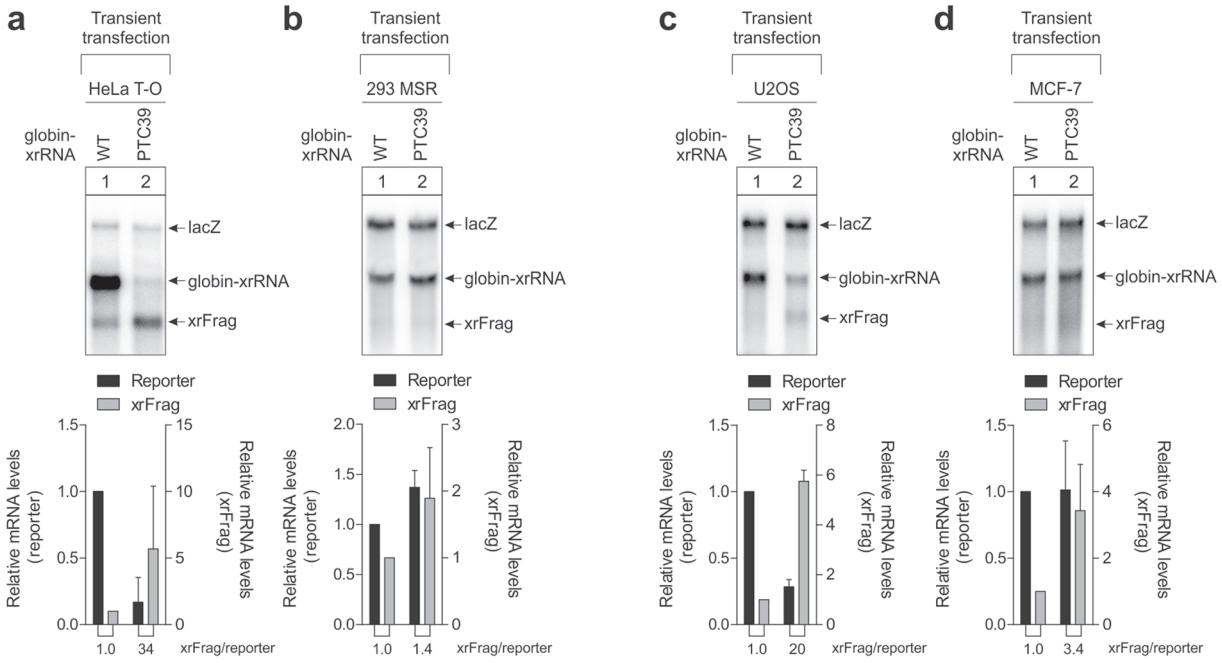


Figure 5. Differential NMD efficiency in transiently transfected human cells. (**a**–**d**) HeLa Tet-Off (HeLa T-O, (**a**), GripTite 293 MSR (**b**), U2OS (**c**) and MCF-7 (**d**) cells were transfected with 0.3 µg plasmid-DNA per 6-well and the RNA was extracted and analyzed by northern blotting. HeLa T-O, 293 MSR and U2OS cells were transfected using calcium phosphate transfection and MCF-7 using jetPRIME transfection reagent. LacZ was co-transfected as a control. The mean values of reporter ± SD (n = 3) were quantified and the PTC-reporter levels normalized to the wild-type control. The xrRNA/reporter ratio is indicated below the graph.

(Threshold cycle) value was calculated. Globin mRNAs levels were normalized to mVenus mRNA levels. For each primer, a 2-fold standard dilution was performed to calculate the primer efficiencies, which ranged between 90–100%. The fold changes were calculated using the $\Delta\Delta C_t$ method[16]. The mean fold changes were calculated from three biologically independent experiments. Primer sequences were: 5′-CAGGCTGCTGGTGGTCTAC-3′ and 5′-CGTGCAGCTTGTCACAGTG-3′ (globin mRNA); 5′-CCATCTTCTTCAAGGACGAC-3′ and 5′-TGATATAGACGTTGTGGCTG-3′ (mVenus mRNA).

Data availability statement. The authors declare that the data supporting the findings of this study are available within the paper.

References

1. Doma, M. K. & Parker, R. RNA quality control in eukaryotes. *Cell* **131**, 660–668, doi:10.1016/j.cell.2007.10.041 (2007).
2. Fatscher, T., Boehm, V. & Gehring, N. H. Mechanism, factors, and physiological role of nonsense-mediated mRNA decay. *Cell Mol Life Sci* **72**, 4523–4544, doi:10.1007/s00018-015-2017-9 (2015).
3. Kervestin, S. & Jacobson, A. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* **13**, 700–712, doi:10.1038/nrm3454 (2012).
4. Holbrook, J. A., Neu-Yilik, G., Hentze, M. W. & Kulozik, A. E. Nonsense-mediated decay approaches the clinic. *Nat Genet* **36**, 801–808, doi:10.1038/ng1403 (2004).
5. Mort, M., Ivanov, D., Cooper, D. N. & Chuzhanova, N. A. A meta-analysis of nonsense mutations causing human genetic disease. *Hum Mutat* **29**, 1037–1047, doi:10.1002/humu.20763 (2008).
6. Schweingruber, C., Rufener, S. C., Zund, D., Yamashita, A. & Muhlemann, O. Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim Biophys Acta* **1829**, 612–623, doi:10.1016/j.bbagrm.2013.02.005 (2013).
7. Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol* **16**, 665–677, doi:10.1038/nrm4063 (2015).
8. Boehm, V., Gerbracht, J. V., Marx, M. C. & Gehring, N. H. Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences. *Nat Commun* **7**, 13691, doi:10.1038/ncomms13691 (2016).
9. Viegas, M. H., Gehring, N. H., Breit, S., Hentze, M. W. & Kulozik, A. E. The abundance of RNPS1, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway. *Nucleic Acids Res* **35**, 4542–4551, doi:10.1093/nar/gkm461 (2007).
10. Linde, L. *et al.* Nonsense-mediated mRNA decay affects nonsense transcript levels and governs response of cystic fibrosis patients to gentamicin. *J Clin Invest* **117**, 683–692, doi:10.1172/JCI28523 (2007).
11. Zetoune, A. B. *et al.* Comparison of nonsense-mediated mRNA decay efficiency in various murine tissues. *BMC Genet* **9**, 83, doi:10.1186/1471-2156-9-83 (2008).
12. Gardner, L. B. Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis. *Mol Cancer Res* **8**, 295–308, doi:10.1158/1541-7786.MCR-09-0502 (2010).
13. Mino, T. *et al.* Regnase-1 and Roquin Regulate a Common Element in Inflammatory mRNAs by Spatiotemporally Distinct Mechanisms. *Cell* **161**, 1058–1073, doi:10.1016/j.cell.2015.04.029 (2015).

14. Boehm, V., Haberman, N., Ottens, F., Ule, J. & Gehring, N. H. 3′ UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep* **9**, 555–568, doi:10.1016/j.celrep.2014.09.012 (2014).
15. Mansour, F. H. & Pestov, D. G. Separation of long RNA by agarose-formaldehyde gel electrophoresis. *Anal Biochem* **441**, 18–20, doi:10.1016/j.ab.2013.06.008 (2013).
16. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101–1108 (2008).

Acknowledgements

We thank Juliane Hancke for technical assistance; the Leptin lab for sharing equipment; Elena Dobrikova and Matthias Gromeier for establishing and Matthias Hentze for sharing the HeLa Flp-In T-REx cells; Kay Hofmann for sharing the U2OS and MCF-7 cells and members of the Gehring lab for useful discussions. This research was funded by a grant from the Deutsche Forschungsgemeinschaft (GE2014/4-1) to N.H.G.

Author Contributions

N.H.G., V.B. and J.V.G. designed the study. J.V.G. performed the experiments. J.V.G. and N.H.G. wrote the manuscript. All authors discussed the results and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-10847-4

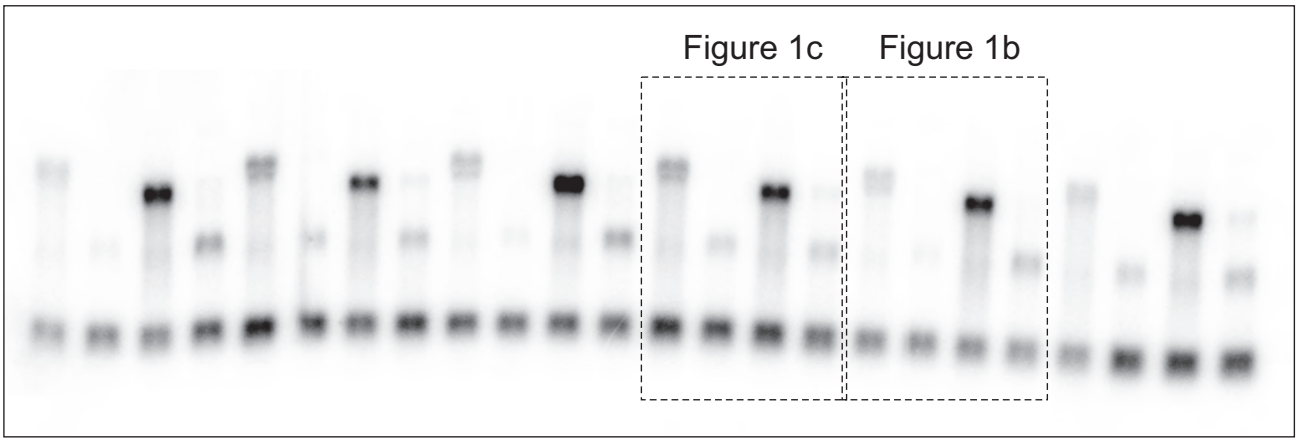
Competing Interests: The authors declare that they have no competing interests.

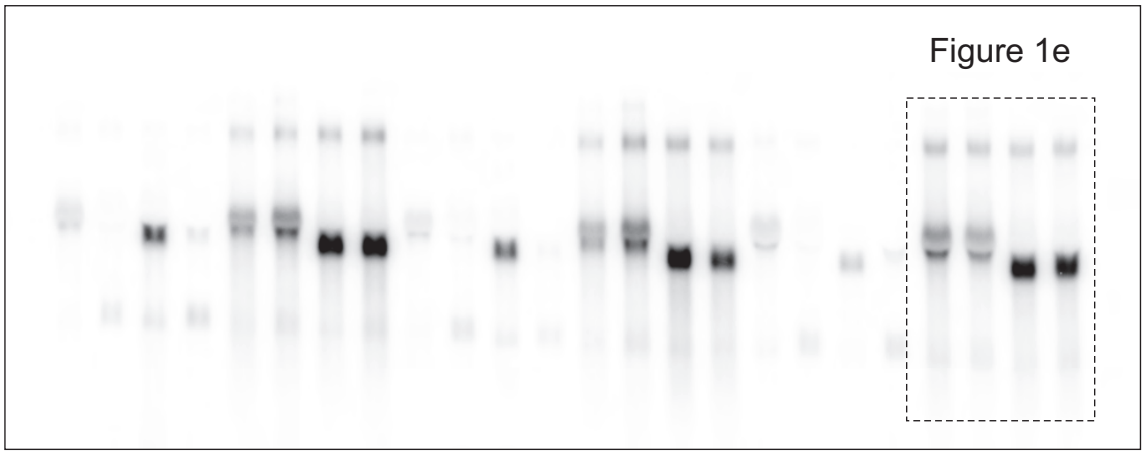
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

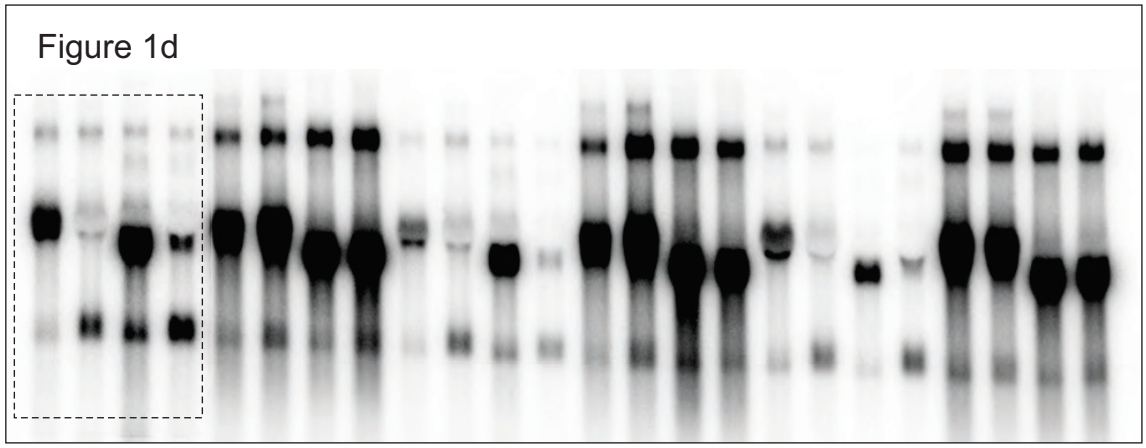
Supplementary Information

Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells

Jennifer V. Gerbracht, Volker Boehm and Niels H. Gehring

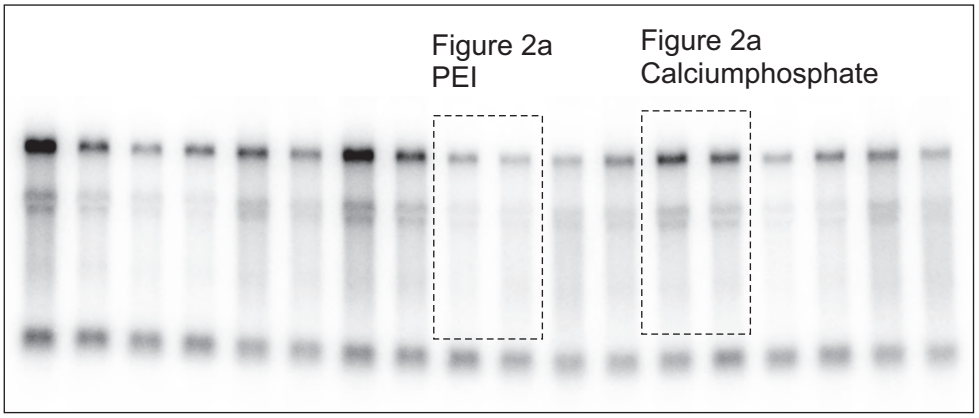


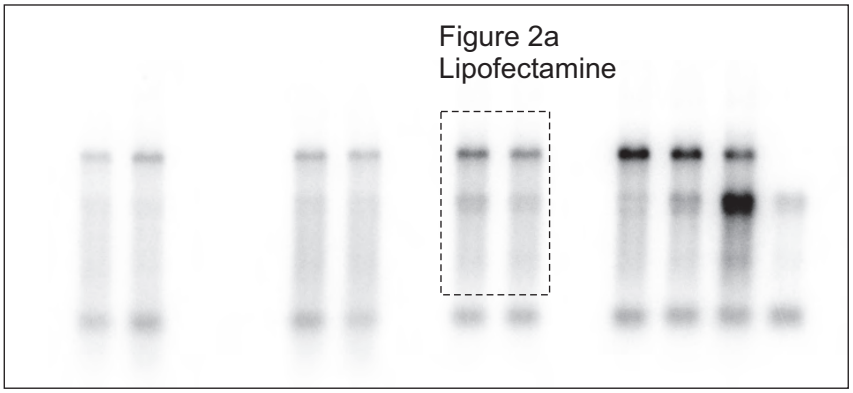


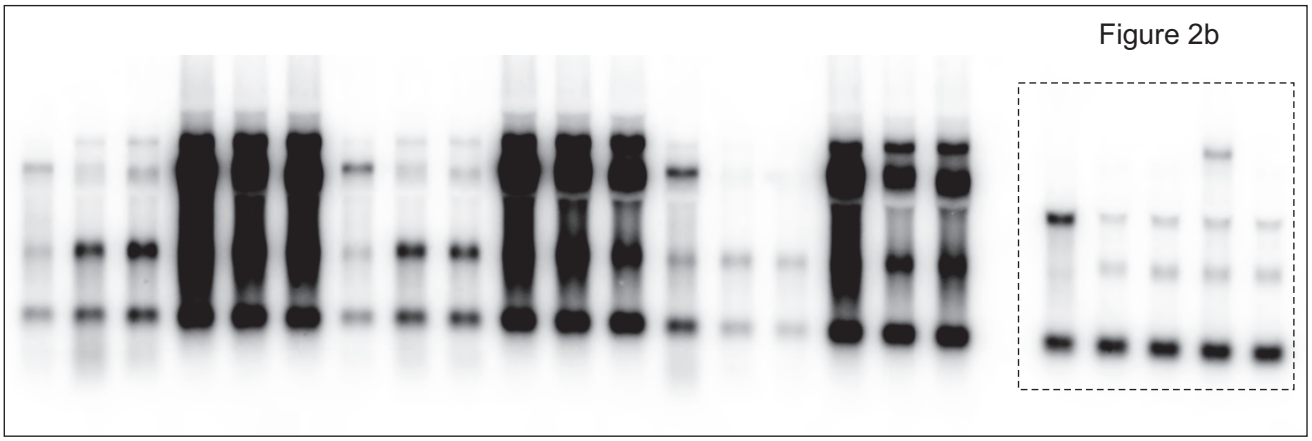


Supplementary Figure 1

Uncropped blots of Figures 1-5 as indicated.







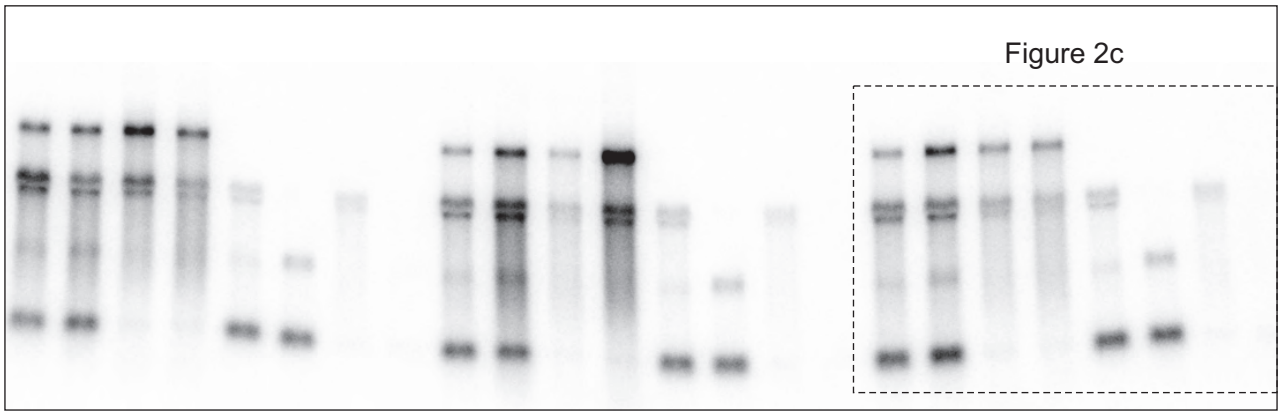
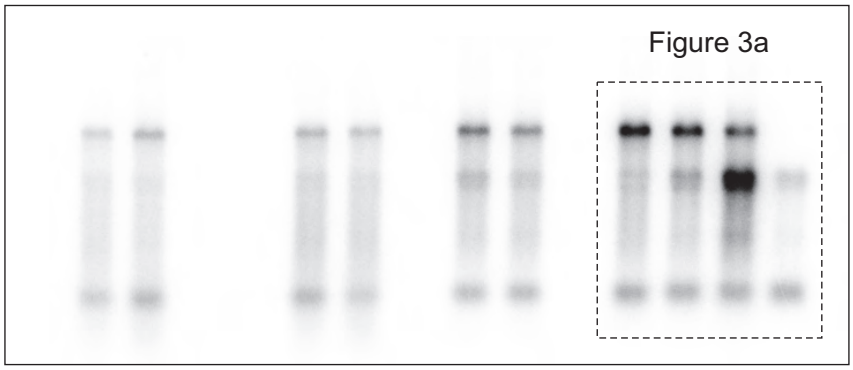
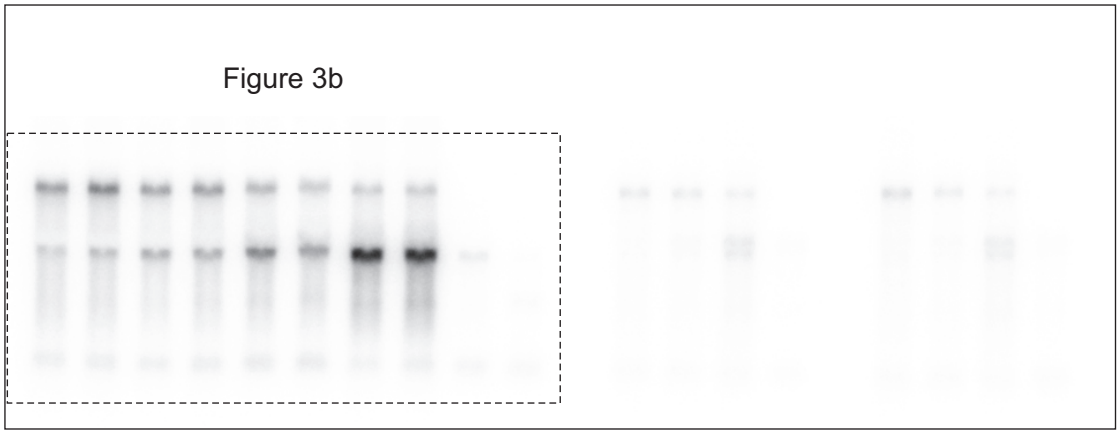


Figure 2c

Supplementary Figure 1, continued

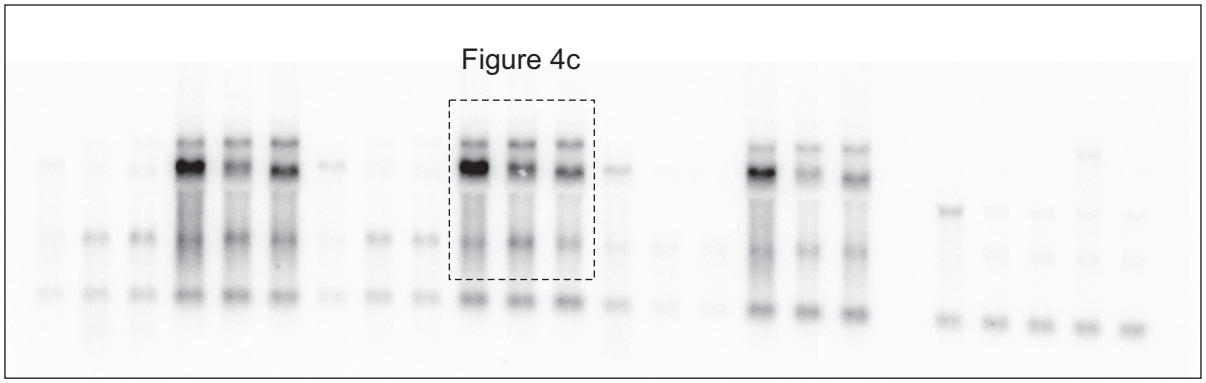
Uncropped blots of Figures 1-5 as indicated.





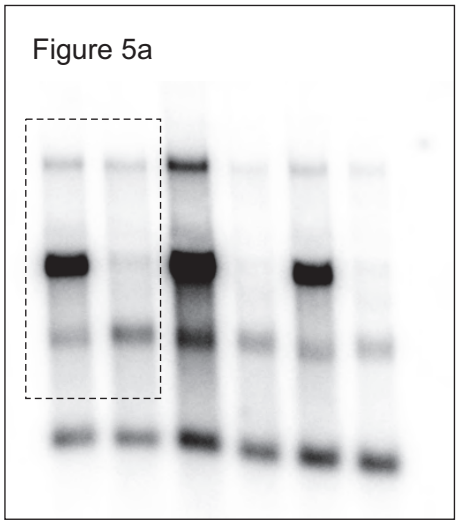
Supplementary Figure 1, continued

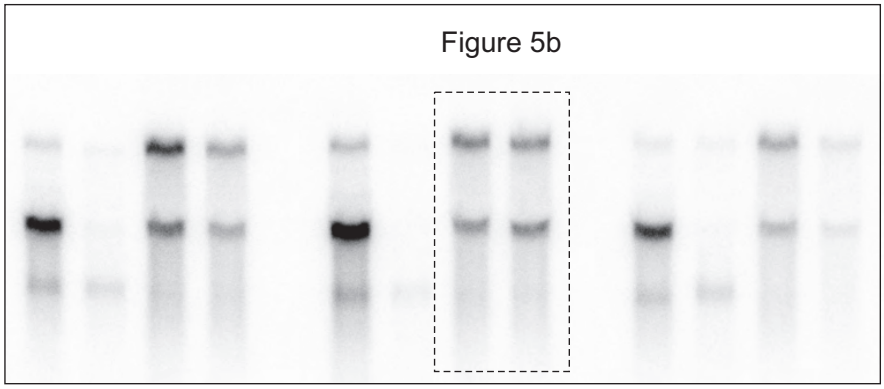
Uncropped blots of Figures 1-5 as indicated.

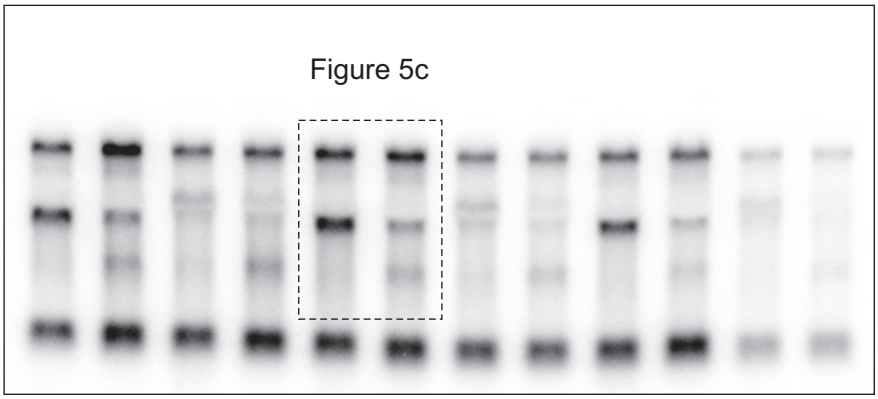


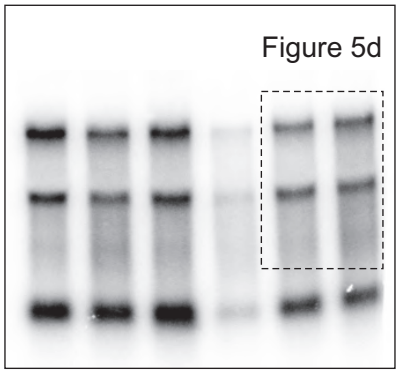
Supplementary Figure 1, continued

Uncropped blots of Figures 1-5 as indicated.









Supplementary Figure 1, continued

Uncropped blots of Figures 1-5 as indicated.

Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts

Author's version of accepted and peer-reviewed manuscript.
Voigt, F., Gerbracht, J.V., Boehm, V., Horvathova, I., Eglinger, J., Chao, J.A. and Gehring, N.H. (2019) Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts. *Nature protocols*, **14**, 1603-1633.

Authors

Franka Voigt[1*], Jennifer V. Gerbracht[2*], Volker Boehm[2], Ivana Horvathova[1,3], Jan Eglinger[1], Jeffrey A. Chao[1,#], Niels H. Gehring[2,#]

[1] Friedrich Miescher Institute for Biomedical Research, CH-4058 Basel, Switzerland.

[2] Institute for Genetics, University of Cologne, D-50674 Cologne, Germany

[3] University of Basel, CH-4003 Basel, Switzerland.

∗ These authors contributed equally to this work.
Corresponding author

Dr. Jeffrey A. Chao

phone: +41.61.697.5173

fax: +41.61.697.3976

e-mail: jeffrey.chao@fmi.ch

ORCID ID: 0000-0002-5895-9156

office: Maulbeerstrasse 66, CH – 4058 Basel

Dr. Niels Gehring

phone: +49.221.470.3873

fax: +49.221.470.5264

e-mail: ngehring@uni-koeln.de

ORCID ID: 0000-0001-7792-1164

office: Zuelpicher Str. 47a, D-50674 Koeln

KEYWORDS RNA decay, RNA degradation, xrRNA, NMD, XRN1, single-molecule fluorescence microscopy, single-molecule imaging, single-particle tracking, single-molecule microscopy, RNA biology, mRNA, northern blot, qPCR, quantitative PCR, fluorescent in situ hybridization, FISH, smFISH, single-molecule FISH,

EDITORIAL SUMMARY This protocol describes how to detect RNA decay intermediates in human cultured cells. Reporter mRNAs can be detected and quantified with molecular biology (northern blot, qPCR) and single-molecule fluorescence microscopy (smFISH, live imaging) approaches.

TWEET A new protocol for detecting and quantifying RNA decay intermediates using molecular biology and single-molecule imaging approaches.

COVER TEASER Detection of RNA decay intermediates

Up to three primary research articles where the protocol has been used and/or developed:

1 Boehm, V., Gerbracht, J. V., Marx, M. C. & Gehring, N. H. Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences. *Nat Commun* 7, 13691, doi:10.1038/ncomms13691 (2016).
2 Gerbracht, J. V., Boehm, V. & Gehring, N. H. Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells. *Sci Rep* 7, 10616, doi:10.1038/s41598-017-10847-4 (2017).
3 Horvathova, I. *et al.* The Dynamics of mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells. *Mol Cell* 68, 615-625 e619, doi:10.1016/j.molcel.2017.09.030 (2017).

Abstract

RNA degradation ensures appropriate levels of mRNA transcripts within cells and eliminates aberrant RNAs. Detailed studies of RNA degradation dynamics have been prevented by the inherent instability of degradation intermediates due to the high processivity of the enzymes involved. In order to visualize decay intermediates and to characterize the spatiotemporal dynamics of mRNA decay, we have developed a set of methods that apply XRN1-resistant RNA sequences (xrRNAs) to protect mRNA transcripts from 5'-3' exonucleolytic digestion. This approach is the only method that can detect the directionality of mRNA degradation and allows tracking of degradation products in unperturbed cells. Here, we provide detailed procedures for xrRNA reporter design, transfection and cell line generation. We explain how to extract xrRNA reporter mRNAs from mammalian cells and their detection and quantification using northern blotting and quantitative PCR. The procedure further focuses on how to detect and quantify intact reporter mRNAs and XRN1-resistant degradation intermediates using single-molecule fluorescent microscopy. It provides detailed instructions for sample preparation and image acquisition using fixed as well as living cells. The procedure puts special emphasis on detailed descriptions of high-throughput image analysis pipelines, which are provided along with the manuscript and were designed to perform spot co-localization, detection efficiency normalization and the quality control steps necessary for result interpretation. The analysis software published here aims to enable non-expert readers to detect and quantify RNA decay intermediates within 4-6 days after reporter mRNA expression.

Introduction

mRNA decay is a highly regulated process that controls gene expression via multiple pathways and through a coordinated series of enzymatic reactions. Accordingly, the regulation of mRNA stability affects cellular mRNA levels during normal development as well as in response to external perturbations[1,2]. It is used for example during the inflammatory response, where cis-acting AU-rich elements (AREs) serve to regulate inflammatory mediators like TNFα via binding of ARE-binding proteins that can increase (HuR) or decrease (e.g. TTP or BRF1) mRNA stability[3,4]. Another class of cis-acting elements termed GREs (GU-rich elements) controls mRNA splicing and decay. GREs are bound by CELF proteins, however GRE-mediated regulation of mRNA decay is not well understood[5]. Interestingly, the conditional regulation of mRNA half-life also applies during the decay of histone mRNAs, which occurs in a cell cycle-dependent manner and serves to coordinate histone protein production with DNA synthesis[6].

Canonical mRNA decay is mediated by 5′-3′ and 3′-5′ exonucleases, which depend on prior deadenylation, decapping or endonucleolytic cleavage steps[7]. For most transcripts, decay is initiated via poly-A tail shortening by either the PAB-specific ribonuclease 2 and 3 complex (Pan2-Pan3), by de-adenylating complexes containing CCR4–NOT subunits or by the poly(A) ribonuclease PARN[8]. mRNA decay proceeds through hydrolyzation of the 5′ cap structure by DCP1-DCP2, which generates 5′ phosphorylated mRNAs that are substrates of 5′-3′ exonucleolytic decay[9,10]. Finally, degradation of the mRNA body is carried out in 5′ to 3′ direction by the cytoplasmic 5′-3′ exoribonuclease 1 (XRN1) and/or with 3′ to 5′ polarity by cytoplasmic exosome components such as RRP44 or EXOSC10[7]. XRN1 has been described to be the main 5′-3′ exonuclease responsible for degradation of cellular mRNAs in the cytoplasm[11]. However, due to its high processivity it has not been possible to directly monitor XRN1-mediated degradation of individual mRNAs in their cellular context. Instead, existing methods had to rely on the stabilization of mRNA decay intermediates via knockdown of degradation machinery components, which directly interfere with the cellular processes involved in mRNA turnover.

To visualize the degradation of unstable mRNA species and to assess the spatial and temporal dynamics of mRNA turnover, we have developed a set of methods that employ flaviviral RNA structures (xrRNAs) to render mRNA reporter transcripts resistant to degradation by XRN1 (Figure 1a). These xrRNA sequences block XRN1-mediated decay by sequestering the 5′ phosphate on the RNA substrate from the nuclease active site[12]. Thus, the stability of otherwise unstable 5′-3′ degradation intermediates is significantly increased and they can be unambiguously distinguished from intact transcripts using northern blotting[13,14], quantitative real-time PCR (qPCR) (Figure 1b) and fluorescent microscopy[15] (Figure 1c).

Applications of the method

The use of xrRNA reporter transcripts enables the assessment of mRNA degradation beyond steady state measurements of mRNA abundance and generates a direct read-out without the need for depletion of the cellular degradation machinery via RNAi knockdowns. The method can be readily implemented in existing workflows since XRN1-resistant decay intermediates are detectable via standard molecular biology methods such as northern blot or qPCR. In combination with state-of-the-art single-molecule fluorescence microscopy techniques, the xrRNA approach provides an easily applicable way to analyze the spatial and temporal dynamics of mRNA degradation in fixed and live cells. The technique is fully compatible with standard mRNA degradation assays such as half-life measurements when factors involved in mRNA degradation are depleted. Additionally, it can be combined with tethering assays to monitor the effect of RNA-binding proteins on transcript stability or with

staining of different subcellular compartments to assess the influence of spatial heterogeneity on mRNA stability.

Modifications in the xrRNA reporter design allow dissection of the contributions of different degradation pathways to the turnover of a transcript (see Experimental Design). Specifically, the interplay of exo- and endonucleolytic degradation during the decay of a target RNA can be assessed by flanking individual regulatory elements with xrRNA elements. We have applied such modifications to study the decay of reporters containing the TNFα or IL6 3′ UTRs and were able to quantify the contributions of different decay pathways[13]. In addition, we have used xrRNA reporter transcripts to investigate the effect of isolated regulatory sequence elements like the 5′-terminal oligopyrimidine element (TOP), AU-rich elements (ARE), and siRNA- or miRNA target sequences. In this context, we have developed a method for real-time observation of Ago2 slicing of reporter mRNAs[15].

xrRNA reporter transcripts can further be applied to study the spatiotemporal dynamics of mRNA turnover using single-molecule methods. So far, we have studied mRNA turnover in cytosolic RNA granules such as P-bodies. Our study revealed, that their degradation occurs independently in the cytosol and that P-bodies are not sites of reporter mRNAs degradation[15].

Our results show that the xrRNA approach represents an ideal starting point to investigate the effect of hitherto uncharacterized mRNA sequence elements or putative mRNA decay factors. With a few experimental steps, the main degradation pathways can be examined and conclusions be drawn about the molecular functions of the respective elements and factors. Since XRN1 is highly conserved across eukaryotic species, the methodology can be applied to a variety of other experimental systems. It is even conceivable that mRNA degradation could be investigated *in vivo* in well-known model organisms such as fruit flies or nematodes using the xrRNA technology.

An attractive goal for the future is to apply the method to study the turnover of endogenous transcripts via integration of xrRNA sequences into genomic loci using CRISPR/Cas9. In addition, it is tempting to combine xrRNA fragments with other single-molecule imaging techniques that can monitor translation via nascent polypeptide labeling to assess the correlation of translation and decay[16-20].

Comparison with other methods

Direct monitoring of mRNA turnover has been difficult due to the instable nature of mRNA decay intermediates and the high processivity of the exonucleases involved. Instead, mRNA degradation rates have been estimated by blocking transcription through application of antibiotics such as actinomycin D, in combination with monitoring the decrease of cellular mRNA levels over time. As long as mRNA decay rates do not change upon addition of actinomycin D, this method can be employed to reliably assess mRNA turnover[21]. However, blocking transcription severely affects cell viability and initiates stress response pathways[22]. Additionally, the global inhibition of transcription also results in the loss of transcripts that encode components of the RNA decay machinery, which prevents accurate measurements of mRNA stability since decay pathways are perturbed.

Another common approach involves the depletion of XRN1 to allow detection of decay intermediates when studying cytosolic mRNA turnover. However, XRN1 interacts with DCP1, a crucial component of the decapping machinery[23]. The presence of XRN1 therefore likely influences the decapping process and its absence affects mRNA decay rates beyond the stabilization of decay intermediates. XRN1 knockdown represents a very intrusive approach to study mRNA turnover while the use of XRN1-resistant fragments within a specific transcript is less invasive and provides an alternative to the depletion of decay pathway components.

In contrast to the methods discussed above, the application of xrRNA reporter constructs allows simultaneous detection of intact mRNAs and degradation intermediates and thereby enables quantification of decay rates. While the insertion of a poly(G) tract allowed detection of decay products in yeast[24], to our knowledge there is no other method that can directly detect decay intermediates in metazoan cells. It is possible, however, to measure degradation rates using mRNA synthesis rates. In order to distinguish newly transcribed mRNAs from the overall cellular mRNA population, several studies have employed metabolic labeling of mRNA using 4-thiouridine. Consequently, metabolic labeling at high temporal resolution followed by precise mRNA quantification and computational modeling has been applied to determine mRNA transcription and decay rates[22]. Since metabolic labeling allows investigation of a multitude of endogenous transcripts without need for genetic manipulation of the target genes, it appears as a cost-effective and minimal interference approach that could be easily combined with the use of xrRNA reporter transcripts to study the dynamics of a regulatory relationship identified via metabolic labeling in more detail.

Limitations

The approach described in this protocol is simple, cost-effective and can be of minimal invasiveness since it provides an immediate read-out for XRN1-mediated 5'-3' decay without the need for manipulation of the cellular degradation machinery. It is limited by the necessity to construct reporter transcripts, which restricts its applicability to endogenous mRNAs or high-throughput approaches. However, we believe that the combination of our method with established genome editing techniques allows the insertion of xrRNA sequences into genomic regions of interest. In this case, however, possible consequences of xrRNA insertions on cellular function should be considered. It is also conceivable that prolonging the half-lives of decay intermediates could have consequences for global gene expression, (e.g. sequestration of critical RNA-binding proteins), however, this is more likely to occur if reporter mRNAs are over-expressed from transient transfection. For live cell imaging, two additional transgenes (PP7 and MS2 coat proteins) need to be introduced into a cell line along with the mRNA reporter. To facilitate experimental set-up, cell lines suitable for single-molecule imaging are available from the Chao lab upon request and additional reagents are available at Addgene.

Experimental design

Here, we describe a set of techniques that apply viral xrRNA elements to block XRN1-mediated 5'-3' decay and generate stable decay intermediates that can be detected by standard RNA molecular biology and imaging techniques. To assay mRNA decay using xrRNA reporter transcripts, researchers need to design reporter as well as control constructs (Step 1), express them either transiently or generate stable cell lines (Steps 2-3), and perform a standard molecular biology (northern blot (Step 18A), qPCR (Step 18B)) or single-molecule imaging experiments (Steps 19-21) as described further below.

Reporter design (Step 1): To function as mRNAs, reporter transcripts need to contain an open reading frame (ORF in graphical overview in Figure 2) that can be translated and – ideally – be used to assess mRNA translation. To this purpose, we commonly include a Renilla luciferase (RL) gene and have utilized a RL activity assay, which is commercially available (Promega) and can be used to quantify protein expression[25]. Alternatively, other ORFs (for instance β-globin) can also be used to generate different reporter constructs. Their translation can be checked by western blot analysis.

To block exonucleolytic digest by XRN1, we incorporated viral XRN1-resistant RNA (xrRNA) fragments downstream of the reporter gene ORF within the 3′ UTR. Many flaviviral xrRNA fragments are structurally well described and we generated reporters featuring elements of the Murray Valley encephalitis (MVE) virus for molecular biology experiments as well as the West Nile virus Kunjin strain (WNV_{KUN}) for single-molecule imaging[12]. These xrRNA elements inhibit XRN1-mediated degradation by the same mechanism and should both be suitable for either molecular biology or single-molecule imaging. They consist of two xrRNA sequences (xrRNA1 and xrRNA2), which block XRN1 by sequestering the RNA 5′ phosphate away from the enzyme[26]. However, XRN1 is not permanently inhibited by xrRNA fragments. Instead, xrRNAs only increase the half-life of the decay intermediates so that these can be detected but will eventually also be degraded. At least one full xrRNA element, consisting of two xrRNA sequences is needed to allow decay intermediate detection[13].

Introducing xrRNA elements in the 3′ UTR could potentially change the stability of the reporter transcript. Therefore, the xrRNA insertion should be designed to not disturb known regulatory regions. In our experience, the position of the xrRNA element relative to the stop codon and polyA site does not substantially influence its efficiency. However, we recommend to include 50-100 nucleotide spacer regions between xrRNA element, stop codon and polyA site. Furthermore, we recommend to generate reporters with and without xrRNA elements to check whether expression and turnover of these constructs are altered.

To allow detection of decay intermediates via northern blotting, it is necessary to design probes that anneal downstream of the xrRNA fragments. Thus, we have incorporated four repeats of the probe-binding sequence in the 3′ UTR of the reporter constructs. Using corresponding probes, XRN1-resistant decay intermediates can be detected as additional, faster-migrating bands during northern blot analysis. For qPCR analysis, we further apply linkers that are ligated to the isolated mRNA molecules. The 5′ cap structure present on intact mRNAs will prevent 5′ linker ligation and therefore linker sequences will allow to specifically detect decay intermediates. Standard qPCR analysis with primers complementary to the reporter gene and linker regions allows quantification of intact and degraded mRNA fractions within one simple and fast experiment. When performing transient transfection, the expression of the reporter mRNAs has to be compared to co-transfected stable mRNAs in order to control for transfection efficiency. Control transcripts used in the anticipated results section of the protocol are plasmids expressing lacZ or mVenus. They can be detected on a northern blot by including a northern probe-binding site. Alternatively, in qPCR analysis the co-transfected control is measured by a specific set of primers. For analysis of stably transfected cell lines, the levels of a constitutively expressed endogenous RNAs are detected to control for total RNA amounts. In the anticipated results section, 7SL RNA levels are detected in northern blots and GAPDH mRNA levels are measured for qPCR experiments.

For single-molecule imaging, xrRNA sequences are flanked with orthogonal PP7 and MS2 stem loop repeat cassettes[27]. Since MS2 stem loops are protected from XRN1-mediated decay by the xrRNA fragments positioned 5′ of the MS2 cassette, this design allows discrimination between intact transcripts (PP7 and MS2) and stabilized degradation intermediates (MS2 only). In fixed cells, this can be done via single-molecule fluorescent in situ hybridization (smFISH) using probes specifically designed to target the reporter gene/PP7 and MS2 repeat regions. In live cells, mRNAs can be detected via fluorescently-labeled coat

proteins that recognize PP7 and MS2 repeats. To this aim, PP7 and MS2 coat proteins (PCPs and MCPs) fused to spectrally distinct fluorescent proteins (e.g. NLS-PCP-GFP and NLS-MCP-Halo) are stably integrated into the HeLa cell genome by lentiviral transduction[28]. In both cases, single-molecule resolution is achieved by recruiting multiple copies of PCPs/MCPs to each RNA molecule as well as by a reduction of the background fluorescence in the cytosol through sequestration of unbound coat proteins to the nucleus via N-terminal nuclear localization signals (NLS). The coat protein constructs mentioned throughout this manuscript are available via Addgene (IDs 104098, 104099).

To benchmark detection efficiencies, it is essential to perform all single-molecule experiments with a control transcript in addition to the reporter construct. Control transcripts should be identical to the mRNA reporter employed but lack the xrRNA fragments between the PP7 and MS2 cassettes. This way, decay intermediates are not stabilized. Control transcripts should always appear as intact dual-labeled particles in an imaging experiment and can be employed to assess detection efficiencies for both fluorescent labels. Reporter and control transcript-encoding plasmids that can be used for preliminary experiments or as cloning templates are available on Addgene (IDs 104095, 104097, 108366 – 108378).

Reporter expression (Steps 2-3): xrRNA reporter mRNA-encoding plasmids as described above are designed for expression in mammalian cell lines. Reporter transcript-encoding plasmids can be transiently transfected or stably integrated into the target cell line (Figure 2). Transient transfection is ideal for quickly obtaining results and allows testing of several different reporter transcripts before starting the laborious and time-consuming generation of a stable cell line.

Of note, transient transfection can result in significant cell-to-cell variations in expression levels, which affects the reproducibility of the performed experiment. For reproducible expression levels, we recommend to use an established human cell line (e.g. HeLa 11ht) that ideally co-expresses a tetracycline transactivator element to induce reporter expression from a tetracycline-inducible locus upon doxycycline addition[29].

During single-molecule imaging experiments, we made similar observations in transient and stably transfected cells. However, transient transfection has been reported to affect the readout of mRNA decay reporter assays during investigation of NMD substrates in certain cell lines[14]. We therefore recommend starting with transient transfections to test the feasibility of an experiment but then to generate stable cell lines to yield robust and reproducible results.

Reporter detection (Steps 3-21): The application of xrRNA reporter transcripts for half-life measurements has an important advantage when compared to standard decay assays: It provides an additional read-out since it allows direct detection of stabilized decay intermediates. Thus, xrRNA reporter transcripts should be applied in qPCR or northern blot experiments to allow rapid and reliable quantification of decay intermediates and to distinguish mRNA degradation from other possible causes of variations in steady-state amounts of an mRNA.

Single-molecule imaging methods can be used to assay the dynamics and spatiotemporal resolution of mRNA decay in single cells. However, while the read-out from an imaging experiment can be very informative, it is also more time-consuming to set up and needs careful calibration before quantitative interpretation of results can be performed. To remedy this need for expert knowledge, the following procedure (Steps 19-21) provides detailed instructions as well as test data and data processing workflows to enable all readers to apply the detection methods presented here.

Materials

Biological materials

- Cell lines for molecular biology analysis: HeLa Tet-Off (Clontech/Takara, cat. no. 631156, RRID:CVCL_V352), Flp-In T-REx 293 (Thermo Fisher Scientific, cat. no. R78007, RRID:CVCL_U427) and HeLa Flp-In T-REx (established by Elena Dobrikova and Matthias Gromeier, Duke University Medical Center).
- Cell line for single-molecule imaging: HeLa 11ht (expressing rtTA2-M2 tetracycline reverse transactivator) generated for Tet-On transcription of a gene of interest inserted into a single locus by flp recombinase[29]
  ! CAUTION Cell lines used in your research should be regularly checked to ensure that they are not infected with mycoplasma.

Reagents

▲CRITICAL Nuclease-free reagents are required for northern blot and quantitative real-time PCR. To avoid contamination and RNA degradation, do not share reagents and kits. Refer to the MSDS before handling toxic or hazardous reagents.

- Mammalian expression vectors for molecular biology analysis: we use the Flp-In T-REx Core Kit for cloning and establishing stable inducible cell lines (Thermo Fisher Scientific, cat. no. K6500-01). Alternatively, for transient transfection use pCI-neo (Promega; cat. no. E1841). Selected reporter plasmids are available on Addgene (IDs 108366 – 108378). Plasmid information is listed in Supplementary Table 1.
- Mammalian expression vectors for single-molecule imaging: mRNA reporter constructs are expressed from custom-made pSF4 vectors, where they are transcribed from a Tet-inducible CMV promoter and flanked by flippase (flp) recognition target (*FRT*) sites for single-locus genomic integration by flp recombinase-mediated cassette exchange (Addgene IDs 104095, 104097).
- Lentiviral expression vectors for coat protein integration needed for live cell imaging: phage plasmids that express protein of interest from a constitutive UbiC promotor (Addgene IDs 104098, 104099). Plasmid information is listed in Supplementary Table 1.
  ! CAUTION Lentiviral particles are a potential biological hazard. All work with lentiviruses must be performed according to biosafety level 2 regulations. Follow your institutional guidelines regarding risk reduction measures and waste disposal.
- Oligonucleotide sequences are listed in Supplementary Table 2.
- DMEM (Dulbecco's Modified Eagle Medium), high glucose with GlutaMAX supplement (Thermo Fisher Scientific, cat. no. 61965-059)
- Fluorobrite™ DMEM imaging medium (Thermo Fisher Scientific, cat. no. A1896701)
- Halo-labeling dye Janelia Fluor 549 (JF_{549})[30] (Tocris, Cat.No.: 6147)
- FBS (Thermo Fisher Scientific, cat. no. 10270-106)
- PBS (Thermo Fisher Scientific, cat. no. 14190-169)
- Doxycycline hyclate (Sigma, cat. no. D9891-1G)
- Hygromycin B Gold (InvivoGen, cat. no. ant-hg-1)
- Plasmid transfection reagent of choice (see step 2)
- peqGOLD TriFast (VWR, cat. no. 30-2010)
  ! CAUTION TriFast is toxic and should be used in a fume hood; wear gloves and a lab coat.
- 1-Bromo-3-chloropropane (Sigma, cat. no. B9673-200ML)
  ! CAUTION 1-Bromo-3-chloropropane is toxic. Handle it inside a fume hood and wear gloves and a lab coat.

- Precipitation Carrier (Molecular Research Center, cat. no. PC 173)
- Ethanol (Carl Roth, cat. no. 9065.4)
  ! CAUTION Ethanol is flammable; keep away from heat, sparks, and open flames.
- 2-Propanol (Carl Roth, cat. no. 6752.4)
  ! CAUTION 2-Propanol is flammable; keep away from heat, sparks, and open flames.
- Ultra pure water (Merck, cat. no. L0015)
- SeaKem LE Agarose (Lonza, cat. no. 50005)
- Formaldehyde solution, ≥37% (wt/vol), non-acidic (Carl Roth, cat. no. P733.1)
  ! CAUTION Formaldehyde solutions and formaldehyde vapors are toxic; prepare the solutions in a fume hood.
- Triethanolamine (Carl Roth, cat. no. 6300.1)
- TRICINE (Carl Roth, cat. no. 6977.3)
- Formamide, deionized (Carl Roth, cat. no. P040.1)
  ! CAUTION Formamide is a toxic material and a teratogen. Handle it inside a fume hood; wear protective gloves and a lab coat.
- Ethylenediamine tetraacetic acid (EDTA; Carl Roth, cat. no. CN06.3)
- Ethidium Bromide (EtBr) 0.07% (wt/vol) solution (AppliChem, cat. no. A2273,0005)
  ! CAUTION Ethidium Bromide is toxic and harmful. Wear gloves and eye glasses when you are using this compound. If skin is exposed to EtBr, extensively wash it with running water.
- Glycerol (Carl Roth, cat. no. 3783.1)
- Bromophenol blue (Sigma, cat. no. B0126-25G)
- ssRNA Ladder (NEB, cat. no. N0362S)
- NaCl (Carl Roth, cat. no. 3957.1)
- tri-Sodium citrate dihydrate (Carl Roth, cat. no. 3580.1)
- Hydrochloric acid (HCl; Carl Roth, cat. no. 4625.1)
- SDS ultra-pure (Carl Roth, cat. no. 2326.3)
  ! CAUTION This compound is harmful. Avoid inhalation and contact with skin and eyes.
- Di-Sodium hydrogen phosphate (Na_2HPO_4; Carl Roth, cat. no. P030.2)
- Sodium hydroxide (NaOH; Carl Roth, cat. no. 6771.1)
  ! CAUTION NaOH is corrosive. Wear gloves and handle under fume hood to avoid eyes and skin damage.
- alpha-P32-Guanosine 5’-triphosphate, 800 Ci/mmol, 10 µCi/µl (α-^{32}P-GTP; Hartmann Analytic, cat. no. FP-808)
  ! CAUTION Follow the applicable local and institutional regulations for handling radioactivity.
- gamma-P32-Adenosine 5'-triphosphate, 800 Ci/mmol, 10 µCi/µl (γ-^{32}P-ATP; Hartmann Analytic, cat. no. FP-801)
  ! CAUTION Follow the applicable local and institutional regulations for handling radioactivity and for the relevant waste disposal policies.
- SP6 RNA Polymerase (NEB, cat. no. M0207L)
- NTP Bundle, 4 x 100 mM (ATP, CTP, GTP, UTP) (Jena Bioscience, cat. no. NU-1014S)
- 10x DTT (NEB, cat. no. B1034A)
- RNAsin, 40 U/µl (Promega, cat. no. N251A)
- DNAse I, RNAse-free, 1 U/µl (Thermo-Fisher, cat. no. EN0525)
- mini Quick Spin RNA Columns (Sigma-Aldrich, cat. no. 000000011814427001)
- T4 Polynucleotide Kinase (10 U/µL) (Thermo-Fisher, cat. no. EK0031)
- T4 RNA Ligase 1 (NEB, cat. no. M0204S)
- Biotin-11-UTP (Jena Bioscience, cat. no. NU-821-BIOX)

- Tris (Carl Roth, cat. no. AE15.2)
- Tween 20 (Sigma-Aldrich, cat. no. P1379
- Albumin Fraction V/Bovine serum albumin (BSA; Carl Roth, cat. no. 8076.2)
- Streptavidin-HRP (Cell Signaling Technology, cat. no. #3999S)
- Western Lightning Plus ECL, Perkin Elmer, cat.no. NEL103E001EA)
- dNTP Mix, Roti®-Mix PCR 3 (Carl-Roth, cat. no. L785.2)
- GoScript Reverse Transcriptase (Promega, cat. no. A5003)
- NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, cat. no. 740609.50)
- GoTaq qPCR Master Mix (Promega, cat. no. A6001)
- Two sets of Stellaris FISH probes (Biosearch Technologies) annealing to the reporter gene or PP7 stem loop cassette (Quasar670) and the MS2 stem loop cassette in the 3′ UTR (Quasar570)[15]
- Paraformaldehyde 20% (wt/vol) aqueous solution (Electron Microscopy Sciences; cat. no. 15713)
  ! CAUTION It is a hazardous cross-linking agent. Wear protective gloves and handle under fume hood.
- Dextran sulfate sodium salt (Sigma Aldrich; cat. no. D6001)
- UltraPure 20x SSC Buffer (Life Technologies; cat. no. 15557036)
- DAPI, dilactate (Sigma; cat. no. D9564)
- Prolong gold antifade reagent (Molecular Probes, Life Technologies; cat. no. P36934)
- 10x PBS pH 7.4 (Thermo Fisher Scientific; cat.no. AM9624)

Equipment

Cell culture

- Nitrile gloves (TouchnTuff 90-600, Ansell)
- Serological pipettes (Greiner Bio-one)
- Fume hood (SafeAir, Hamilton)
- Refrigerated centrifuge (Eppendorf)
- Cell culture hood (Maxisafe 2020, Thermo Scientific)
- Cell culture incubator (CO_2 at 5%, humidified, 37 °C) (Galxy 170R, New Brunswick, Eppendorf)
- Automated cell counter (Biorad, cat. no. 1450102)
- Counting slides (Biorad, cat. no. 1450011)
- Falcon™ 50ml conical centrifuge tubes (Corning, cat. no. 352070)
- Tissue culture–treated 6-well plates (Corning, cat. no. 3506)
- Tissue culture–treated 12-well plates (Corning, cat. no. 3513)

Molecular biology

- Nanodrop 2000 spectrophotometer (Thermo Scientific, cat. no. ND-2000)
- Shaking heat block (e.g. HLC Heating-ThermoMixer)
- Cellulose acetate (CA) 0.2 µm pore size membrane filter (Sartorius, cat. no. 11107--47----ACN)
- Microwave
- Horizontal gel electrophoresis chamber for 20 x 20 cm gels (Carl Roth, Rotiphorese Unit PROfessional IV, cat. no. 2941.1)
- Electrophoresis power supply (e.g. Consort EV243)
- UV gel imaging system (e.g. Bio-Rad Gel Doc XR+ Gel Documentation System)
- Sterile disposable scalpel (e.g. Bayha scalpel-blades + handle)
- Plastic tray (at least 20 x 10 cm)

- Rocking platform (e.g. Biometra WT16 rocking platform)
- Lab dish 35 x 25 x 4 cm (Carl Roth, cat. no. 8455.1)
- Blotting paper, grade 703 (VWR, cat. no. 732-0591)
- Roti-Nylon 0.2 (Carl Roth, cat. no. AE50.1)
- Saran wrap
- Folded paper hand towels
- Bio-Link 254 UV crosslinker (Vilber Lourmat, cat. no. BLX 254)
- Hybridization oven (e.g. Thermo Scientific Hybaid Shake 'n' Stack, VWR cat. no. HYBA6243)
- Hybridization bottle, 250 x 35 mm or 300 x 35 mm (e.g. VWR cat. no. 732-1822 and 732-1823)
- Geiger Counter (e.g. Berthold LB 124 SCINT, cat. no. 43727-10)
- BAS storage phosphor screens, BAS-IP MS 2025 E (VWR, cat. no. 28-9564-75)
- Imaging cassette (e.g. VWR Exposure cassette, cat. no. 29-1755-24)
- Phosphorimager (e.g. GE Healthcare Typhoon FLA 7000)
- Chemiluminescence Imaging System (e.g. Thermo Fisher Scientific myECL Imager)
- ImageQuant TL 8.1 software for quantification (https://www.gelifesciences.com/en/us/shop/protein-analysis/molecular-imaging-for-proteins/imaging-software/imagequant-tl-8-1-p-00110)
- Hard-Shell PCR Plates (Bio-Rad, cat. no. HSL9601)
- Microseal 'B' Adhesive Seals (Bio-Rad, cat. no. MSB1001)
- CFX96 Real-Time PCR Detection System (Bio-Rad, cat. no. 1855195)

smFISH

- Microscopy slides (Menzel Gläser, Thermo Scientific, cat. no. 12164682)
- Microscope cover glass (⌀ 18 mm for 12-well plate) (Marienfeld Laboratory Glassware, Ref: 0111580)
- Stellaris® FISH probes against the gene of interest (LGC Biosearch Technologies); for design instructions see Supplementary Procedure 3.
- Petri dishes 100x15 mm (VWR, cat. no. 10416-320)
- Parafilm (Sigma Aldrich, cat. no. P7793)
- Hybridization incubator at 37˚C (e.g. FisherbrandTM, cat. no. 13-247-30Q)
- Rubber tubing (Masterflex, cat. no. 6402-24 and Fisher, cat. no. Tygon R-3603)
- Filter flask (1 liter, Pyrexplus; Fisher Scientific, cat. no. 10-180G)
- Bottle stopper (Fisherbrand, cat. no. 14-135N)
- A wide-field epifluorescence microscope for fixed cell imaging (see Equipment setup)

Live cell imaging

- A multipoint confocal spinning disk microscope equipped for live cell imaging (see Equipment setup)
- Multi-color calibration slide for channel alignment (e.g. Argolight, type Argo-SLF-001)
- Imaging dish (µ-Dish), 35 mm, high, glass bottom (ibidi, cat. no. 81158)

Image analysis

- Fiji[31] including the TrackMate plugin[32] (https://imagej.net/TrackMate)
- Matlab (Mathworks, https://www.mathworks.com/downloads) running the spot detection algorithm AIRLOCALIZE[33] (http://www.timotheelionnet.net/software/)

- KNIME Analytics Platform (version 3.6.0 or later, https://www.knime.com/downloads)[34] with additional update sites configured (see Equipment setup)
- Decay analysis workflows and test data available from the Chao lab (https://data.fmi.ch/PublicationSupplementRepo/?group=gchao)

Reagent setup

- 20x TT is 0.6 M TRICINE and 0.6 M Triethanolamine. Prepare 1 liter as follows. Add 89.5 g of triethanolamine and 107.5 g of TRICINE and fill up to 900 ml with autoclaved distilled water. Stir until components are dissolved. Bring to 1 liter with autoclaved distilled water and filter the solution using a 0.2 µm membrane filter. The buffer can be stored at room temperature (15-25 °C) for up to a year. Yellow discoloring can occur that from our experience does not affect the buffer's performance.
- 20% (wt/vol) SDS stock solution: Add autoclaved distilled water to 500 g SDS up to a volume of 2.5 liter and stir over night. The stock solution can be stored at room temperature for up to a year.
- RNA sample buffer is 2 µl 0.5 M EDTA pH 8.0, 500 µl deionized Formamide, 35 µl of 37% (wt/vol) Formaldehyde and 50 µl of 20x TT. Add one drop of 0.07% (wt/vol) EtBr solution. The buffer can be stored at -20 °C for up to a year.
- RNA loading buffer is 50% (vol/vol) glycerol, 1 mM EDTA and 0.25% (wt/vol) bromophenol blue. Prepare in autoclaved distilled water. The buffer can be stored at room temperature for up to a year.
- 20x SSC is 3 M NaCl and 0.3 M tri-Sodium citrate dihydrate. Prepare 1 liter as follows. Add 175.3 g of NaCl and 88.2 g of tri-Sodium citrate dihydrate and fill up to 900 ml with distilled water. Adjust the pH to 7.0 with HCl. Autoclave and store at room temperature for up to a year.
- Church buffer is 0.5 M Na_2HPO_4, 1 mM EDTA, 7% (vol/vol) SDS, pH 7.2. To prepare 1 liter, first add 500 ml of autoclaved distilled water to 70.98 g of Na_2HPO_4 and 0.29 g of EDTA and adjust pH to 7.2 with NaOH. Add 350 ml of 20% (wt/vol) SDS solution, mix and fill up to 1 liter with autoclaved distilled water. Store at room temperature for up to a year.
  ▲**CRITICAL** Pre-heat to 65 °C before use or to facilitate proper buffer mixing, prolonged storage at 65 °C is not recommended. Discard when buffer turns yellow.
- Northern wash buffer 1 is 2x SSC and 0.1% (vol/vol) SDS. Prepare 1 liter using 100 ml of 20x SSC and 5 ml of 20% (wt/vol) SDS solution, fill up with autoclaved distilled water. Store at room temperature for up to a year.
- Northern wash buffer 2 is 0.2x SSC and 0.1% (vol/vol) SDS. Prepare 1 liter using 10 ml of 20x SSC and 5 ml of 20% (wt/vol) SDS solution, fill up with autoclaved distilled water. Store at room temperature for up to a year.
- 10x TBS stock is 0.5 M Tris and 1.5 M NaCl. Prepare 1 liter by dissolving 60.5 g Tris and 87.6 g NaCl in 800 ml of distilled water and adjust the pH to 7.5 with HCl. Bring to 1 liter with distilled water. Autoclave and store at 4 °C for up to a year.
- TBS-T is 1x TBS stock solution with 0.2% (vol/vol) Tween 20. Prepare 1 liter by mixing 100 ml of 10x TBS stock with 2 ml of Tween 20 and filling up with autoclaved distilled water. Store at room temperature for up to a year.
- 5% (wt/vol) BSA TBS-T To make 50 ml of this solution, dissolve 2.5 g BSA in 50 ml of TBS-T and mix by very mild stirring to avoid foaming. Prepare freshly before use.

- smFISH fixation solution is 4% (vol/vol) paraformaldehyde in 1x PBS. Prepare 50 ml using 10 ml (one vial) of 20% paraformaldehyde and 5 ml of 10x PBS, fill up with nuclease-free water.
  ▲**CRITICAL** Store at 4 °C for up to one week. Longer storage is not recommended.
  ! CAUTION Formaldehyde is an irritant and should be handled under chemical fume hood.
- smFISH hybridization buffer is 10% (wt/vol) dextran sulfate, 10% (vol/vol) formamide in 2x SSC supplied with working concentration of fluorescent DNA oligomer probes. Prepare 150 µl per 4 coverslips using 75 µl of 20% (wt/vol) dextran sulfate, 15 µl formamide, 15 µl 20x SSC and 45 µl nuclease-free water. We recommend preparing 20% (wt/vol) dextran sulfate in nuclease-free water at least one day in advance as dissolution of the dextran sulfate powder may take O/N incubation at 37°C while shaking.
  ! CAUTION Formamide is an irritant and teratogen that is easily absorbed through the skin and should be opened under chemical fume hood when equilibrated to room temperature.
- FISH wash buffer is 10% (vol/vol) formamide in 20x SSC. Prepare 50 ml using 5 ml of formamide, 5 ml of 20x SSC and 40ml nuclease-free water.
  ▲**CRITICAL** Store at 4 °C for up to one week. Longer storage is not recommended.
- DAPI staining solution is 500 ng/ml of DAPI in 1x PBS. Prepare 100 ml using 50 mg DAPI dissolved first in 70 ml of nuclease-free water. Then add 10 ml of 10x PBS, filled up with nuclease-free water. The solution is stable over months at 4 °C.
  ▲**CRITICAL** DAPI does not dissolve readily in PBS.

Equipment setup

- For assembly of a table-top aspirator to remove supernatant from coverslips adhere to instructions given by Trcek et al. (2012)[35]. Briefly, place a rubber stopper into a filter flask, attach Masterflex tubing to the outlet of the rubber stopper and Tygon tubing to the flask outlet itself. Attach a Pasteur pipette to the end of the Masterflex tubing and connect the Tygon tubing with the vacuum outlet in the laboratory.
- For fixed cell imaging, we use a Zeiss Z1 wide-field microscope equipped with an AxioCam 506 mono camera and an X-Cite 120 EXFO metal halide light source for fluorescence, a halogen light source for transmitted light and a Plan-APOCHROMAT 100x 1.4 NA oil objective.
- For live cell imaging, we use an inverted confocal microscope (Olympus IX81) that is equipped with a CSU-X1 scanhead (Yokogawa) and two precisely-aligned back-illuminated EvolveDelta EMCCD cameras (Photometrics). Images are collected using a 100x 1.45NA PlanApo TIRFM oil immersion objective (Olympus), green (Semrock, FF01-617/73-25) and red (Semrock, FF02-525/40-25) emission filters and solid-state lasers (491 nm and 561 nm Cobolt) as excitation sources. The microscope is equipped with an incubation chamber that provides heating and CO_2 regulation. Before starting image acquisition, equilibrate the incubation chamber around the microscope to 37 °C and 5% CO_2. Align cameras using a multi-color calibration slide (e.g. Argo-SLF-001).
- For data analysis, install the KNIME Analytics Platform (version 3.6.0 or later)[34] and configure additional update sites (*File > Preferences > Install/Update > Available Software Sites* (using the **Add...** button) to add the FMI software site (https://community.knime.org/download/ch.fmi.knime.plugins.update/). Then add the FMI KNIME Plugins via *Help > Install New Software…* and selecting *FMI KNIME Plugins* from the newly added update site. For additional information concerning

installation, refer to the “readme.txt” file provided along with the software and test data.

Procedure

Design of reporter mRNA ● TIMING ~ 1-5 d

1. Generate a reporter transcript according to the instructions provided above in 'Experimental design'. For initial tests or standard assays, use available reporter plasmids from the Chao and Gehring labs (Addgene IDs 104095, 104097, 108366-108367 or 108375-108376 (see Reagents)). Alternatively, use standard molecular cloning approaches to obtain appropriate reporter constructs suitable for either transient transfection or stable cell line generation.

Expression of reporter mRNA

2. For expression of reporter mRNAs using either transient transfection, follow option A. For stable integration, follow option B. See Experimental design for a discussion of advantages and disadvantages of both methods.

A Transient transfection ● **TIMING ~2 d**

(i) Seed cells of choice into 6-well plates at confluencies appropriate for the transfection reagent used and grow them overnight in medium appropriate for the cell line being used (e.g., for HeLa cells, use DMEM with 10% (vol/vol) FBS).

▲**CRITICAL STEP** Prepare 1-2 wells for each experimental sample. Although 1 well yields typically enough material for all downstream steps (depending on the transfection efficiency), the extra well can be used for western blot or other control analyses.

(ii) Transfect the cells 16-24 h after seeding with the transfection reagent of choice (may require medium change prior to transfection). We routinely perform transient transfections using the calcium phosphate method, however other transfection reagents have been tested yielding similar results[14]. The amount of transfected plasmid can be adjusted depending on the expected readout, however 0.1-1 µg per well works well with most reagents and applications. If possible, include a transfection control, e.g. a plasmid encoding for a fluorescent protein in the same well as the experiment.

(iii)16-24 h after transfection, inspect transfection efficiency (if controls were included) and continue the protocol with successfully transfected samples (step 3). If longer expression times are required, change medium and grow cells up to additional 24 h.

B Stable integration ● **TIMING ~14-28 d**

(i) To generate stable cell lines, use a specialized genomic insertion system (e.g. Flp-In T-REx) and follow the manufacturer's instructions. Commercially available or custom-made cell lines can be used for the molecular biology protocol (e.g. Flp-In T-REx 293 from ThermoFisher or HeLa Flp-In T-REx). For imaging experiments, we apply a HeLa 11ht cell line that allows expression of reporter or control transcripts from a single tetracycline-regulated locus for transgene expression after recombinase-mediated cassette exchange[29].

Preparing samples for northern blotting and qPCR ● TIMING **2-3d**

▲**CRITICAL** The following sections (Steps 3-18) describe how to detect decay intermediates via northern blotting and qPCR. For instructions on how to perform single-molecule imaging experiments directly proceed to step 19.

▲**CRITICAL** When performing northern blotting or qPCR using transient transfection, skip this step and proceed to step 4.

3. For stable cell lines, seed cells in 6-well plates at 40% confluency (approximately 0.3 x 10^6 cells) and grow them overnight. If inducible expression systems were used, induce expression of the reporter with the recommended concentration of inducing agent.
   ▲**CRITICAL STEP** For the Flp-In™ T-REx 293 Cell Line and HeLa Flp-In T-REx cell line, use up to 1 µg/ml doxycycline for efficient induction. Maximum expression levels of reporter mRNA are typically observed at around 8 h after induction, although expression can in general be performed for up to 24 h.

RNA extraction ● TIMING 2h

4. Wash cells with 1 ml of PBS
5. Harvest cells with peqGOLD TriFast using 1 ml of reagent per 6-well plate well according to the manufacturer's instructions. Transfer to new 1.5-ml microcentrifuge tube.
   ▲**CRITICAL STEP** After addition of TriFast, rapidly pipet each sample up and down multiple times in order to reduce the viscosity.
6. Incubate the samples for 5 min at room temperature.
   ■ **PAUSE POINT** The TriFast samples can be stored for several weeks at -20 °C.
7. Add 150 µl of 1-Bromo-3-chloropropane to each sample, mix vigorously by shaking and incubate for 10 min at room temperature.
8. Centrifuge the samples at 12,000g for 10 min at 4 °C.
9. Take 450 µl of the upper aqueous phase and transfer to new 1.5-ml microcentrifuge tube.
   ▲**CRITICAL STEP** Do not transfer or touch the interphase in order to minimize DNA contamination.
10. Add 450 µl of 2-Propanol, vortex and incubate for at least 10 min at 4 °C.
    ▲**CRITICAL STEP** If low yield of RNA is expected, add 1-2 µl of Precipitation Carrier to each sample.
11. Centrifuge the samples at 12,000g for 10 min at 4 °C.
12. Aspirate the supernatant and wash the pellet with 1 ml of 75% (vol/vol) Ethanol.
    ? TROUBLESHOOTING
13. Centrifuge the samples at 12,000g for 10 min at 4 °C.
14. Repeat step 12-13.
15. Aspirate supernatant completely and briefly dry the pellet for 5 min by leaving the lid of the 1.5-ml microcentrifuge tube open.
16. Add 20 µl ultrapure water to the pellet and incubate for 10 min at 50-65 °C on a shaking heat block at 500-900 rpm to help complete resuspension of the RNA.
17. Quantify RNA concentration using a Nanodrop 2000 spectrophotometer.
    ▲**CRITICAL STEP** The concentration should be in the range of 500-2000 ng/µl.
    ? TROUBLESHOOTING
    ■ **PAUSE POINT** Purified RNA samples can be stored for several months at -20 °C.

Detection of RNA degradation using northern blotting and qPCR

18. For northern blotting follow option A, for quantitative RT-PCR follow option B (see Experimental Design for details).

A Northern blotting ● TIMING 2-3d

i. ***Gel electrophoresis (Steps i-viii):*** To prepare the denaturing 1% (wt/vol) agarose gel for electrophoretic separation of RNA samples, boil 2.6 g agarose in 230 ml autoclaved distilled water.
▲**CRITICAL STEP** The amounts given are for 20 x 20 cm gels, for smaller gels adjust the amounts accordingly.

ii. Under a fume hood add 12.5 ml 20x TT buffer and 10 ml of freshly filtered 37% (wt/vol) formaldehyde. Mix well, pour gel in the appropriate tray with the desired combs and let the gel solidify for 30 min.

iii. Transfer the gel to the electrophoresis chamber filled with 1x TT buffer and equilibrate the gel for at least 10 min.
▲**CRITICAL STEP** To minimize run artifacts, fill the chamber before casting the gel and pre-cool the chamber to 4 °C (either on ice or in the cold room).

iv. For each experimental sample, add 16 µl of RNA sample buffer to a new 1.5-ml microcentrifuge tube.
▲**CRITICAL STEP** Prepare a sample with RNA marker mixed with RNA sample buffer to allow RNA size determination.

v. Add 2-5 µg of total RNA to the respective 1.5-ml microcentrifuge tube containing the RNA sample buffer.
▲**CRITICAL STEP** The volume of added RNA solution should not exceed 4 µl, otherwise increase the volume of the RNA sample buffer accordingly.

vi. Incubate the samples for 15 min at 65 °C on a heat block. Cool briefly for 2 min at 4 °C.

vii. Add 2 µl of RNA loading buffer to each sample and mix well.

viii. Load the gel with the samples and run at 160 V for approximately 2-4 h, depending on the gel size. Stop the run when the bromophenol blue running front is almost at the end of the gel.
? TROUBLESHOOTING

ix. ***Capillary transfer (Steps ix-xxiii):*** Take a picture of the gel using a UV-transilluminator. (Figure 3)

x. Cut the gel with a clean scalpel according to the expected size of the transcript of interest and control RNA. For example, when using the β-globin reporter which generates RNA fragments in a range of 1 – 2 kb and the 7SL RNA control with a size of 0.3 kb, we recommend cutting the gel above the 28S rRNA and below the bromophenol blue running front.
? TROUBLESHOOTING

xi. Transfer the gel to a suitable plastic tray and wash for 5 min with sufficient amount of autoclaved distilled water (approximately 50 ml) on a rocking platform.

xii. Bridge a lab dish (35 x 25 cm) with the gel tray (20 x 20 cm).

xiii. Cover the tray with blotting paper (30 x 20 cm).

xiv. Fill the lab dish and wet the blotting paper with 10x SSC, remove all air bubbles between the paper and the tray with a sterile serological pipette.

xv. Place the gel upside down on the blotting paper, remove bubbles between gel and blotting paper.
▲**CRITICAL STEP** Using the smooth bottom side of the gel for contact with the membrane prevents transfer artifacts

xvi. Cover the remaining surface of the blotting paper and the lab dish with saran wrap to prevent bypass transfer of 10x SSC.

xvii. Wet nylon transfer membrane (approximately 10 x 20 cm) in autoclaved distilled water and place on top of the gel, remove air bubbles.

xviii. Add one piece of wet blotting paper (10 x 20 cm), soaked in autoclaved distilled water, remove bubbles.

xix. Add one piece of dry blotting paper (10 x 20 cm), remove bubbles.

xx. Place a stack of 5 – 10 cm of folded paper hand towels on top of the blotting stack. Ensure that the lowermost paper has no creases, as this will affect the transfer.

xxi. Add a plastic tray on top of the blotting stack to distribute the weight and add about 0.5 – 1 kg of weight (e.g. two filled 500 ml glass bottles).

xxii. Transfer for 4h or overnight.

xxiii. Disassemble the transfer setup and immobilize the transferred RNA on the nylon membrane by UV crosslinking (254 nm, 0.12 J/cm^2) from both sides.

? TROUBLESHOOTING

■ **PAUSE POINT** The dry membrane can be stored in a sealed plastic bag for several weeks at -20 °C.

xxiv. ***Hybridization (Steps xxiv-xxxiv):*** Prehybridize the nylon membrane for 1-2 h in 25 ml of Church buffer while rotating at 65 °C in a hybridization oven.

xxv. Discard the prehybridization solution and add 20-25 ml of Church buffer containing the radioactive northern blot probes (e.g. β-globin reporter probe and 7SL control probe). For instructions on how to radioactively label probes, see instructions in Supplementary Procedure 1.

▲**CRITICAL STEP** We recommend using radioactively labeled RNA probes to achieve the highest sensitivity. However, working with radioactive material requires access to isotope working facilities and appropriate waste disposal. A non-radioactive alternative is to use biotin-linked nucleotides for RNA labeling which can be detected by chemiluminescence using peroxidase-coupled streptavidin. Their use is described in Supplementary Procedure 2. Potentially, Supplementary Procedure 2 can also be adapted to use digoxigenin labeled or CLICK-functionalized nucleotides.

Incubate over night while rotating at 65 °C in a hybridization oven.

xxvi. Collect the hybridization solution for potential re-use (may exhibit weaker signals, do not use more than twice).

xxvii. Wash the membrane for 15 min with 20-30 ml of northern wash buffer 1 while rotating at 65 °C in a hybridization oven.

xxviii. Discard the washing solution and repeat step xxvii.

xxix. Discard the washing solution and wash the membrane for 15 min with 20-30 ml of northern wash buffer 2 while rotating at 65 °C in a hybridization oven.

xxx. Discard the washing solution and repeat step xxix.

xxxi. Discard the washing solution, take out the membrane and dry the membrane on blotting paper for 30 min.

xxxii. Wrap the membrane in saran wrap, place the membrane in an imaging cassette and lay the BAS storage phosphor screen on top. Expose for 4h – 24h.

■ **PAUSE POINT** The hybridized membrane can be stored for a few days at room temperature. However, the signal will fade due to the short half-life of P32.

xxxiii. Detect the signal using a phosphorimager.

? TROUBLESHOOTING

xxxiv. ***Quantification (Step xxxiv):*** Use the ImageQuant TL 8.1 software or similar software for quantification (Figure 4). For transiently transfected cells, quantify the intensity of full-length reporter RNA, xrRNA decay intermediate and the co-transfected control for each experimental sample. Calculate the full-length reporter/co-transfection

control and xrRNA decay intermediate/co-transfection control ratios and compare to control samples (e.g. control knockdown). For stable transfections, quantify the intensity of full-length reporter RNA, xrRNA decay intermediate and the 7SL control for each experimental sample. Calculate the full-length reporter/7SL control and xrRNA decay intermediate/7SL control ratios and compare to control samples (e.g. control knockdown).

B Quantitative real-time PCR ● TIMING 2d

▲**CRITICAL** The mRNA levels of the full-length reporter and the XRN1-resistant decay intermediate are measured relative to an internal control (Figure 5). In transiently transfected samples the internal control is a co-transfected construct. In stably transfected samples, any housekeeping gene can be used.

i. To ligate an RNA-linker to RNA fragments with a 5' monophosphate, mix the following reagents in a nuclease free 1.5 ml tube by pipetting. Incubate at room temperature for 2h, then over night at 18 °C.

Reagents	Volume (µl) per reaction
RNA linker (100 µM)	2
Buffer T4 RNA Ligase (10x)	3
RNasin (40 U/µl)	0.5
ATP (10 mM)	3
15 µg total RNA	x
H_2O	Up to 30
T4 RNA Ligase I (10 U/µl)	1
Total reaction volume	30

ii. Add the reagents listed in the table below to the linker ligation reaction mix from Step i in order to perform reverse transcription. Mix well by pipetting and incubate for 2h at 42 °C in a benchtop heater.

Reagents	Volume (µl) per reaction
Linker ligation reaction from step i	30
VNN-oligo dT20 (100 µM)	3
dNTPs (10 mM each)	3
GoScript Buffer (5x)	12
$MgCl_2$ (25 mM)	9
RNasin (40 U/µl)	1.5
GoScript Reverse Transcriptase (160 U/µl)	1.5
Total reaction volume	60

iii. Purify the cDNA using the NucleoSpin Gel and PCR Clean-up kit according to the manufacturer's instructions. Elute in 50 µl of the provided NE buffer. Dilute 1:5 (vol/vol) in H_2O to use in qPCR.

iv. To amplify the full-length reporter and XRN1-resistent fragment with qPCR, prepare the mix listed below in 96-well plates. For each condition prepare 3 wells (technical triplicate).
▲**CRITICAL STEP** Exact pipetting is required to minimize variation between triplicates. Change tips after each pipetting step to ensure equal amounts of reagents in each well.

Reagents	Volume (µl) per reaction
GoTaq qPCR Master Mix (2x)	5
Sense Primer (10 µM)	0.5
Antisense Primer (10 µM)	0.5
cDNA (1:5 dilution)	4
Total reaction volume	10

v. Perform qPCR by placing the 96-well plate in a qPCR machine and follow the program described below.

	Step	Temperature (°C)	Duration (min)	Cycles
1	Denaturing	95	3:00	
2	Denaturing	95	0:15	40
3	Hybridization + Plate Read	60	0:30	
4	Elongation	72	0:10	
5	Melt Curve + Plate Read	65 to 95	Increment 0.5 °C/0:05	

vi. After the run, check the melting curve for non-specific amplification. Only one peak should be visible in the melting curve. To confirm, run the qPCR samples on an agarose gel.

vii. ***Quantification (Steps vii-x):*** Calculate the fold changes according to the comparative C_T method[36]. In brief: Calculate the mean of the technical triplicates (C_T mean).

viii. Deduct the corresponding control (e.g. co-transfected control, housekeeping gene) from the C_T mean for each target (reporter or decay fragment): ΔC_T.

ix. For each target, deduct the control (e.g. wild type) from the test condition (e.g. premature stop codon-containing construct): $\Delta\Delta C_T$.

x. Calculate the fold change as $2^{-\Delta\Delta CT}$. This value is the fold change of the mRNA in the test condition compared to the control.

Preparing single-molecule Imaging experiments ● TIMING 2-30d

▲**CRITICAL** Single-molecule imaging of mRNA decay intermediates can be performed in fixed cells using fluorescent in situ hybridization (smFISH, Step 21 Option A, Figure 6a) or via direct observation of labeled reporter transcripts in living cells (Step 21 Option B, Figure 6b). For correct interpretation of the imaging data derived from both techniques, first provide the controls described in this section.

19. For benchmarking detection efficiencies, prepare a co-localization control cell line analogous to the HeLa 11ht cell line expressing a transcript that is identical to the reporter transcript applied for the assay but lacks the xrRNA elements (for information on reporter

design see Figure 2 or Experimental design). Alternatively, a control cell line can be obtained from the Chao lab upon request.
▲**CRITICAL STEP** Preparation of a control cell line with comparable fluorescence levels for benchmarking of detection efficiencies is crucial for correct quantification of degradation intermediates.

20. Characterize the reporter gene expression profile over time via luciferase activity assays using cells harvested at different time points after induction of reporter gene expression as described by Horvathova et al. (2017)[15]. Alternatively, characterize particle densities in the cytosol via imaging trials at different time points after induction to ensure that mRNA densities are not too high for accurate single-particle tracking and image analysis. To determine optimal particle densities test spot detection and tracking (steps vi-xiii) on images obtained on different time points after induction. Only choose such experimental conditions that allow accurate tracking of all detected particles. For the HeLa cell lines and reporters described here, this should result in not more than a few hundred particles per cell.

Detecting RNA degradation using single molecule imaging

21. Perform the imaging experiments as described in Option A (fixed cells) and Option B (live cells). For both (reporter and control) cell lines, collect data on at least three separate occasions imaging 10-50 cells each time.

A Fluorescent in situ hybridization ● TIMING 5d

i. ***Sample preparation (Steps i-xii):*** Seed 20,000 cells/well in 1 ml DMEM + 10% (vol/vol) FBS + 1% (vol/vol) Pen/Strep onto glass coverslips placed into a 12-well plate. Make sure the cells are spread evenly and incubate cells for ~48h.
▲**CRITICAL STEP** Shorter or longer growth times are possible but make sure that cells have attached on the coverslips to enable imaging of cytoplasmic transcripts. Do not let cells become confluent.

ii. After 48h, induce expression of mRNA reporters by addition of doxycycline for 2h.
▲**CRITICAL STEP** Shorter or longer expression times are possible depending on the experiment performed (see step 20). Make sure that particle densities in cells are not so high that they prevent accurate spot detection.

iii. After 2h, aspirate media and wash 2x with 1x PBS to remove residual doxycycline-containing media.

iv. Either directly fix cells by addition of 1 ml fixation solution per well or re-fill wells with 1 ml DMEM + 10% (vol/vol) FBS + 1% (vol/vol) Pen/Strep and fix at later time points. After addition of fixation solution, incubate 10 min at RT while gently rocking. Aspirate fixation solution and wash cells 2x with 1x PBS.
▲**CRITICAL STEP** Fixation at different time points after transcription shut-off allows quantification of decay dynamics over time. Adjust time points to half-live of intact reporter transcripts and stabilized decay intermediates as e.g. determined by qPCR.

v. For permeabilization of the cell membrane, apply 1 ml 70% (vol/vol) Ethanol per well and incubate at 4 °C for at least 1 hour.
■ **PAUSE POINT** Incubation period may be O/N or even extended to several days.

vi. Aspirate supernatant and wash cells 2x with 1x PBS. Add 1 ml of wash buffer per well and incubate 5 min while gently rocking. Repeat this washing step once again.

▲**CRITICAL STEP** Do not aspirate wash buffer after the second wash. Make sure that coverslips do not dry out.

vii. Add 150 nM of each FISH probe set (each targeting a region upstream and downstream of the xrRNA sequences) to complete the hybridization solution. For information on how to design FISH probes see Supplementary Procedure 3. Then incubate coverslips with hybridization solution by dropping 35 µl onto the bottom of a Petri dish and gently lowering the coverslip (cells facing downwards) into the hybridization solution. Shield coverslips from light and incubate the Petri dishes at 37 °C for at least 4 hours to O/N. Protect coverslips from drying out by providing water reservoirs (e.g. 50 ml Falcon tube caps taped to the dish bottom) and by sealing Petri dishes with Parafilm.

▲**CRITICAL STEP** Make sure that the hybridization solution distributes over the whole surface of the coverslip and no air bubbles remain. More than one coverslip can by placed into a single Petri dish.

viii. Provide 1 ml pre-warmed wash buffer in each well of a fresh 12-well plate and place coverslips carefully inside (cells facing upwards). Wash the coverslips for 30 min at room temperature while rocking the plate. Protect from light. Repeat the 30 min washing step once.

ix. Wash coverslips twice with 1x PBS.

x. Aspirate PBS and add 1 ml DAPI solution per well to stain cell nuclei. After a few seconds, aspirate the DAPI solution and wash 2x with 1x PBS. Leave coverslips in 1x PBS to prevent drying out.

xi. Apply a drop of Prolong Gold mounting medium onto a microscopy glass slide. Take a coverslip out of the 12-well plate and remove excessive PBS by gently touching the edge of the coverslip with a tissue. Invert the coverslip and place it upside down onto the mounting media drop. Store coverslips protected from light O/N to let the medium solidify.

xii. After ~12h, seal coverslips with transparent nail polish and store slides at 4 °C until imaging.

■ **PAUSE POINT** The coverslips can be stored in the dark at 4 °C for up to 6 months.

xiii. ***Image acquisition (Steps xiii-xiv):*** For image acquisition, use the DAPI channel to identify suitable (non-overlapped) cells and avoid bias.

xiv. Acquire optical sections using 240 nm z-steps, spanning a 5 µm z-depth. Typical exposure time for Quasar570 is 1s and for Quasar670 is 2s. 10-50 ms are used to acquire each plane in the DAPI channel.

▲**CRITICAL STEP** Make sure to use exposure times that do not saturate but are well within the dynamic range of the camera.

? TROUBLESHOOTING

xv. ***Spot detection (Steps xv-xxii):*** Select a representative number of z-stacks for further analysis. For training purposes, compare to images provided under "Data1/FISH_TestData/1-RawData" for exemplary experimental data and under "Data1/FISH_ControlData/1-RawData" for co-localization control data to benchmark detection efficiencies.

xvi. Make sure that the two RNA channels of each z-stack are precisely aligned. Correct offset between channels by using e.g. the "Translate" command in the image processing software Fiji[31]. Use the Search Bar (pressing L) in Fiji to find and run the Fiji commands mentioned throughout this protocol.

xvii. Select a single ROI containing a single cell using the rectangular selection tool on all channels simultaneously. Prune three-channel z-stacks to match ROI size using the "Crop" command. Before cropping, we recommend rotating the image around the

position of a particular cell using the "Rotate" command to avoid including parts of neighboring cells and background artifacts.

xviii. Divide channels into individual z-stacks using the "Split Channels" command.

xix. Project z-stacks onto a single *x*,*y* plane by using the "Z Project" command and choosing "Maximum Intensity" as projection type. Make sure to only apply the maximum projection to in-focus images and project the same number of images for all acquired z-stacks.

xx. Save projected images for all channels via "Save As > Tiff…". Make sure that individual channels are named consistently to allow identification during batch processing in AIRLOCALIZE[33]. Compare to test/control data provided in the "2-ProcessedData" folders.

xxi. Perform spot detection to sub-pixel resolution in AIRLOCALIZE as described previously[37].

▲CRITICAL STEP Before applying the batch analysis to data from a particular channel and experiment, define intensity thresholds based on the signal-to-noise ratios (SNRs) in individual images from that data set and by manually inspecting spot detection results of those images.

? TROUBLESHOOTING

xxii. Save coordinate files (.loc3) next to image files (.tif) for each channel in a parental folder that will serve as input directory for further processing in the KNIME[34] workflow. Compare spot detection results to test and control results provided in the "3-AirlocalizeOutput" folders.

xxiii. ***Spot co-localization and data analysis (Steps xxiii-xxxii):*** Load the spot co-localization workflow called "Segmentation, Spot Colocalization and Detection Efficiency Correction" (Software1) available from the Chao lab website (https://data.fmi.ch/PublicationSupplementRepo/?group=gchao) in KNIME[34]. Make sure that node #IDs are displayed by clicking on the "Append the IDs to node names" icon in the symbol bar.

▲CRITICAL STEP mRNA turnover can be monitored via the detection of two orthogonal fluorescent labels on individual mRNA transcripts. In order to distinguish between intact mRNAs and stabilized decay intermediates in a fixed cell experiment, it is therefore necessary to co-localize the signal in the two RNA channels. For correct interpretation of the co-localization data, it is further necessary to take detection efficiencies into account and use them to correct the final particle numbers. Detection efficiencies can be determined via co-localization of control data from a cell line that expresses a reporter that is always double-labeled because it cannot produce stabilized decay intermediates. We have designed a KNIME[34] workflow that incorporates all of the above analysis steps and allows rapid data correction and interpretation. In the following section (Steps xxiv-xxxii), we describe how to use this "Segmentation, Spot Colocalization and Detection Efficiency Correction" pipeline step-by-step.

xxiv. Configure and execute the "Configuration" node (# 94):

- "Input Folder": Choose the parent folder that contains the processed data along with the AIRLOCALIZE output. For the training data provided here, choose "Data1/FISH_TestData/3-AirlocalizeOutput".
- "Name of the first/second spot channel": Enter a name that allows identification of images from different channels (default Cy3/5).
- "Name of the segmentation channel": Enter a name that allows identification of images from the channel used for nuclear/cytoplasmic segmentation (default DAPI).

- "Input Folder Control": Choose the parental folder that contains the processed control data along with the AIRLOCALIZE output. For the control data provided here, choose "Data1/FISH_controlData/3-AirlocalizeOutput".

xxv. The "Set distance threshold (px)" node (# 42) allows adaptation of the distance used to define a spot pair as co-localizing (default = 5 pixel). Make sure the threshold matches your image parameters. For definition of a useful distance thresholds look at merged and aligned (multi-channel) example images: Magnify until individual pixels are visible and count distance (px) between spots belonging to the same channel in close proximity. The threshold has to be chosen such that spots originating from the same particle in different channels will be identified as co-localizing while neighboring spots in the same channel will not.

xxvi. Check that all data were loaded correctly: Execute and view the "Interactive Segmentation View" nodes (# 848 and # 141 for control and test data respectively) in the orange data visualization annotation. Scroll through the table to check that correct image series were loaded. Open the interactive view panel by clicking on an image in the labeling column and switch to the "Bounding Box Renderer" to check visually that spot positions and coordinates match.

? TROUBLESHOOTING

xxvii. "Calculate detection efficiencies" based on spot co-localization of control data by executing node # 854. The branch leading to this node performs pairwise spot co-localization by first measuring the distances of all spots in the first channel to all spots in the second channel. It then defines co-localizing pairs of spots by recursively selecting mutual nearest neighbors and classifies spot pairs as co-localizing if their distance is below the cut-off defined in step xxv. All not co-localizing spots are defined as orphans. Last, the branch calculates detection efficiencies ($e_{C1,C2}$ in Box 1) from the fractions of co-localizing and orphan spots in each cell[15].

▲CRITICAL STEP Inspect the output table to evaluate co-localization results. Detection efficiencies for each channel ($e_{C1,C2}$) are depicted in the last two table columns. Make sure that they are as high as possible ($e_{C1,C2} \approx 0.9$). Since in a control experiment only dual-labeled particles exist, the vast majority of particles should be co-localizing and only very few orphans should be detected.

? TROUBLESHOOTING

xxviii. Execute and inspect the output tables generated by node # 881, which activates the second branch and performs the co-localization analysis for the test data. It will then "Calculate statistics and correct for detection efficiencies" obtained in step xxvii.

xxix. Execute and inspect the output table generated by node # 886. The branch leading to this node calculates a summary of all output data from all cells and filters relevant results. The table lists (i) the "Mean(Fraction of colocalizing particles outside)", which corresponds to the average fraction of intact reporter transcripts in the cytoplasm; (ii) the "Standard deviation(Fraction of colocalizing particles outside)" corresponding to the standard deviation of the average given in (i); (iii) "Count(Cell)", which is the number of cells included in the analysis; (iv) and (v) the "Mean(Detection efficiency (C1 or C2))", that was obtained from the control data and used to correct particle numbers in both channels (C1 and C2).

xxx. Execute the "XLS Sheet Appender" node (# 887) to output all results as a single "Results.xlsx" file within the input directory.

xxxi. Inspect the "selected summary" tab in the output results file. The table lists the same output that was explained in step xxix. The main experimental result is the "Mean(Fraction of colocalizing particles outside)", which corresponds to the average fraction of intact reporter transcripts in the cytoplasm.

xxxii. For training, compare results with the output provided in the "4-KNIMEOutput" folder.

B Live cell imaging ● TIMING 3-4d

▲**CRITICAL** Live-cell imaging experiments require co-expression of fluorescent coat proteins for labeling of mRNA transcripts. To this aim, stably integrate NLS-MCP-Halo and NLS-PCP-GFP into the reporter and control cell lines via lentiviral transduction. Use FACS to select cells with low expression levels of both coat proteins to minimize background fluorescence. For a discussion of different fluorescent proteins and their integration using lentiviruses refer to Halstead et al. (2016)[28]. Alternatively, parental 11ht cell lines[29] that co-express MCP and PCP but do not yet contain reporter transcripts are available from the Chao lab upon request. Lentiviral transfer plasmids encoding NLS-PCP-GFP and NLS-MCP-Halo are available from Addgene (IDs 104098, 104099). Transient expression of coat proteins in preliminary experiments is also possible but involves parallel uptake of three different plasmids (reporter construct + two orthogonal coat proteins). Consequently, identification of cells that express all transfected constructs at low fluorescence levels suitable for live cell imaging can be experimentally challenging.

i. ***Sample preparation (Steps i-iii):*** Seed 30,000 cells in 2 ml DMEM + 10% (vol/vol) FBS + 1% (vol/vol) Pen/Strep into a 35 mm glass bottom imaging dish (ibidi) approximately 48 h prior to the experiment.
▲**CRITICAL STEP** Take care that cells are homogeneously distributed within the dish to allow uniform growth and improve cytoplasmic imaging of transcripts.

ii. Incubate cells for 2 days at 37 °C and 5% CO_2.

iii. Label Halo-tags by incubation with 1 ml 100 nM JF_{549} [30] in pre-warmed DMEM + 10% (vol/vol) FBS. After 30 min at 37 °C and 5% CO_2 remove medium, wash 3 times with pre-warmed PBS and add pre-warmed Fluorobrite™ DMEM + 10% (vol/vol) FBS + 1 µg/ml doxycycline to induce mRNA expression.
? TROUBLESHOOTING

iv. ***Image acquisition (Steps iv-v):*** Use the red laser (NLS-MCP-Halo channel) at low intensities to identify cells for image acquisition

v. Collect 10-50 frame image series per region of interest (ROI) in a single plane and acquire image series in streaming mode in both channels simultaneously using laser powers and camera gain that achieve maximum SNR without saturating the camera. In addition, laser powers and exposure time should be adjusted to limit bleaching of the sample.
? TROUBLESHOOTING
▲**CRITICAL STEP** For accurate single-particle tracking, exposure times should be no longer than 50 ms to ensure that fast particles can be unambiguously identified in between frames.

vi. ***Single-particle tracking (Steps vi-xiv):*** Select a representative number of image series obtained from three independent experiments for further analysis. Pick images of healthy cells with high SNRs and representative particle densities. For training purposes, compare to images provided under Data2/Live_TestData/1-RawData for exemplary experimental data and under Data2/Live_ControlData/1-RawData for co-localization control data needed to benchmark detection efficiencies.
▲**CRITICAL STEP** To monitor the turnover of individual transcripts, it is necessary to reconstruct particle movement over time via SPT using two spectrally distinct fluorescent labels. For unbiased co-localization of the two labels, it is essential to

determine particle trajectories in both channels individually. Therefore, perform steps vi-xiii independently for each channel.

▲**CRITICAL STEP** Detailed step-by-step instructions on how to perform single-particle tracking (SPT) of reporter mRNAs in TrackMate[32] (Fiji) have been published before, see Voigt et al. (2018)[38]. Thus, in the following (steps vii-xiii) we provide only basic instructions that enable the reader to proceed with the analysis.

vii. Make sure that the two channels of each image series are precisely aligned. Correct offset between channels by using e.g. the "Translate" command in the image processing software Fiji[31].

viii. Select a single ROI using the freehand selection tool. Use the "Restore Selection" command to apply the same ROI to both channels. Compare to the processed training data in "2-ProcessedData".

▲**CRITICAL STEP** For accurate SPT, select only those areas of the cells that exhibit high quality SNR without introducing bias with respect to subcellular localization.

ix. Use the "Properties" command to calibrate global image parameters.

x. Decide how many frames to include in the analysis and, if needed, reduce the length of the image series accordingly using the "Slice Keeper" function. Accurate tracking for co-localization analysis is best performed on a small number of frames (typically 3-5) since this limits the number of data points that has to be visually inspected.

xi. Launch the Fiji plugin TrackMate[32] while having the first image series selected.

xii. SPT as performed by TrackMate consists of (i) spot detection and (ii) the linking of particle positions in consecutive frames into tracks (tracking). Detect spots using the "Laplacian of Gaussian" (LoG) detector and optimized detector settings for the detection of single mRNA particles. Use the "Simple LAP tracker" as particle-linking algorithm and set maximum linking and gap closing distances to match physiological diffusion coefficients. Perform tracking but do not filter tracks as this can be done within the co-localization pipeline. For more detailed instructions on how to perform SPT of mRNA transcripts in TrackMate see Voigt et al. (2018)[38].

xiii. Export tracking results via the "Analysis" button in the "Display Options" interface. Save the resulting "Spots in tracks statistics" table as an .xls file with exactly the same name as the input image series. Compare output to the training data output provided in "3-TrackMateOutput".

xiv. Repeat steps x-xiii for the second channel image series with identical tracking parameters.

xv. ***Identification of decay intermediates via track co-localization (Steps xv-xxv):*** For data preparation, assemble the two .xls files exported after SPT along with the image series used for tracking in a single folder for each cell. Combine the folders containing tracking and imaging data for each cell in one or several parent folders. Repeat the same for the co-localization control data.

▲**CRITICAL STEP** In order to distinguish between intact mRNAs and stabilized decay intermediates in a live-cell experiment, it is necessary to co-localize the tracks derived for both labels using SPT. For correct interpretation of the co-localization data, it is further necessary to take detection efficiencies into account and use them to correct the final particle numbers. Detection efficiencies can be determined via co-localization of control data from a cell line that expresses a reporter that must always be double-labeled and cannot exhibit stabilized decay intermediates. We have designed a KNIME[34] workflow that incorporates all of the above analysis steps and allows rapid data correction and interpretation. In the following (Steps xvi-xxv), we describe how to use this pipeline step-by-step.

▲**CRITICAL STEP** Make sure that names of image- and corresponding coordinate files

are identical except for the file extension.

xvi. Load the workflow called "Track-based colocalization analysis (with recursive track pairing and particle number correction)" (Software2) available from the Chao lab website (https://data.fmi.ch/PublicationSupplementRepo/?group=gchao) in KNIME[34]. Make sure that node #IDs are displayed by clicking on the "Append the IDs to node names" icon in the symbol bar.

xvii. Configure and execute the "Data input" node (# 391):

- "Input directory": Choose the parent folder that contains the combined tracking and imaging file folders from several cells and experiments (for training data provided choose: Data2/Live_TestData/3-TrackMateOutput).
- "File extension": Enter the file extension of the coordinate files (xls) derived from SPT.
- "Channel 1 and 2 identifier": Choose identifiers that distinguish input file names belonging to both channels.

▲**CRITICAL STEP** Naming depends on the microscope/software used for data acquisition. Here, default is set to "Red/Green" but can be any description of a wavelength. Choose part of the image file name that allows unambiguous identification of each channel.

- "Load images for quality control". Tick this option if you want to load image files into the pipeline. Identify image files by providing the "image file extension" (e.g. tif). Loading image files will increase computational time but is preferable in the beginning for quality control reasons.
- "Distance cut-off (µm)": Set the distance cut-off that defines a spot pair as co-localizing within the same reference frame is set to default = 0.3 µm.

▲**CRITICAL STEP** Do not modify this parameter in the beginning. Once estimates for co-localization rates have been obtained, co-localization can be made more stringent by reducing this value.

- "Pixel size (µm)": Enter the pixel-size to calibrate imaging data. Default is set to 0.09 µm but depends on microscope set-up.
- "Input directory for control data": Choose this option to identify the parental folder that contains the combined data of the co-localization control (for training data provided choose: Data2/Live_ControlData/3-TrackMateOutput).

xviii. Configure "Track parameters" in node # 405: Choose the "Minimum track length" (default = 3 frames) and the "Minimum number of co-localizations" needed to classify a group of tracks as co-localizing (default = 2 events). This value should be adjusted if longer trajectories are used for analysis. Execute node # 405.

xix. To check that all data were loaded correctly, execute and view the "Interactive Segmentation View" node (# 454) in the orange data visualization annotation. Scroll through table to check that correct image series were loaded. Open the interactive view panel by double-clicking on an image in the labeling column and switch to the "Bounding Box Renderer" to visually check that spot positions (on images) and coordinates (provided by TrackMate output) match. **? TROUBLESHOOTING**

xx. "Calculate detection efficiencies" based on co-localization of control data by executing node # 864. For detailed description of track co-localization see Voigt et al. (2018)[38]. Briefly, the branch leading to this node performs pairwise spot co-localization by first measuring the distances of all spots in the first channel to all spots in the second channel. It then defines co-localizing pairs of spots by recursively selecting mutual nearest neighbors and classifies spot pairs as co-localizing if their distance is below the cut-off defined in step xvii. Next, tracks are classified as co-

localizing if they contain a minimum number of co-localizing spot pairs (user input defined in step xviii, default = 2). All tracks that were not classified as co-localizing are called orphans and usually represent stabilized decay intermediates. If they are shorter than the minimum track length defined in step xviii (default 3 frames) they are excluded from the analysis. Last, the branch calculates detection efficiencies ($e_{C1,C2}$ in Box 1) from the fractions of co-localizing and orphan tracks in each cell[15].

▲**CRITICAL STEP** Inspect the output table to evaluate co-localization results. Table columns depict: (a) the input file path; (b) the number of co-localizing tracks for each channel (n_{Coloc} in Box 1); (c) the number of orphan tracks for each channel ($n_{C1,C2}$); (d) the total number of tracks (= co-localizing + orphan) for each channel; (e) the total number of particles detected per cell ($N = n_{C1} + n_{C2} + n_{Coloc}$); and (f) detection efficiencies for each channel ($e_{C1,C2}$). Make sure that detection efficiencies are as high as possible ($e_{C1,C2} \approx 0.9$). Since in a control experiment only dual-labeled particles exist, the majority of particles should be co-localizing and only very few orphans should be detected.

? TROUBLESHOOTING

xxi. Execute and inspect the interactive table generated by the "JavaScript Table View" node (# 426). It contains the results of the co-localization analysis performed with the experimental data.

xxii. Execute and inspect the interactive table generated by the "JavaScript Table View" node (# 439). The table lists original as well as corrected particle numbers per cell. Table columns depict: (a-e) as described in step xx; (f) the corrected number of co-localizing tracks for each channel (m_{Coloc}); (g) the corrected number of orphan tracks for each channel ($m_{C1,C2}$); and (h) the corrected fraction of co-localizing particles (f_{Coloc}), which corresponds to the fraction of intact mRNA transcripts out of all particles detected. For details on how corrections are calculated, see Box 1.

xxiii. In particular, inspect the corrected number of channel 2 orphans. For the experimental set-up described above, it corresponds to the corrected number of mRNA that are labeled with NLS-PCP-GFP only.

▲**CRITICAL STEP** Since such transcripts should not exist under the described experimental set-up, this value allows assessment of the co-localization control cell line for detection efficiency calculation and should be close to 0. If values are significantly below 0, SNR for the control cell line is worse than for the experimental cell lines and detection efficiencies are over-correcting the true particle numbers. If values are significantly higher than 0, then particle numbers might have been under-corrected.

? TROUBLESHOOTING

xxiv. Execute the "XLS Sheet Appender" node (# 866) to output all results as a single "Results.xlsx" file within the input directory.

xxv. Inspect the output in the "Results.xlsx" file. In particular look at the "Summary – Mean" tab: it lists the mean fraction of co-localizing particles (corrected for detection efficiency), which corresponds to the average fraction of intact particles observed per cell. The output further provides the total number of particles included in the analysis ("# Particles") and the number of cells analysed ("# Cells"). Compare results to the output provided for the training data in "4-KNIMEOutput".

Troubleshooting

Troubleshooting advice can be found in Table 1.

Table 1: Troubleshooting Table

Step	Problem	Possible reason	Solution
12	No visible pellet	Low amount of RNA in samples	Seed more cells. Add Precipitation carrier to step 8.
17	Low RNA yield	Pellet lost during purification	Add Precipitation carrier to step 8. Carefully aspirate supernatants, visually inspect pellet at all times.
18A viii	Bromophenol blue running front lost – 7SL control detection impossible	Gel ran too long or at too high voltage	Run another gel with fresh samples and stop the gel run earlier.
18A ix	Smeared lanes	RNA degradation	Avoid contamination with RNAses, obtain new RNA samples. Clean gel chamber and prepare fresh running buffer.
18A xxiii	Inefficient/incomplete transfer	Insufficient buffer volume or incorrect transfer setup	Repeat the transfer procedure with a new gel, provide enough buffer and adhere to the setup instructions. Specifically, avoid buffer shortcuts from the reservoir to the paper stack. Check successful transfer by UV illumination.
18A xxxiii	Faint, faded signal on parts of the blot	Space between Phosphor screen and the saran-wrapped membrane	Smoothen the phosphor screen on the membrane to ensure close contact. Also, remove any bubbles in the saran wrap.
21A xiv	Autofluorescence	Old or stressed cells	Always use fresh cells. Avoid using coverslips with cells grown to saturation.
	Low signal, very low background	Incomplete permeabilization of cell membrane	Increase incubation time with permeabilization solution.

	Too much background signal	Probe hybridization is unspecific	Decrease hybridization time. Reduce the amount of probe used.
21A xxii	AIRLOCALIZE over-or undercounts; Unresolvable RNA spots	Intensity thresholds are too high or low; too many particles per cell	Manually optimize thresholds determined for individual conditions; reduce induction time.
21B iv	Suboptimal labeling	Free fluorescent dye remaining in cells; labeling time too short	Increase the number of wash steps after fluorescent dye labeling. Increase labeling time. Use fresh dye aliquots.
	Autofluorescence	Labeling time too long; old or stressed cells	Reduce labeling time to below 30 minutes. Always use fresh cells. Avoid using dishes with cells grown to saturation.
21B v	Bleaching of fluorescent signal	Limited photostability of fluorescent dyes; formation of radical oxygen species	Reduce total irradiation by adaptation of laser intensities and exposure time. Use specific mounting medium designed to increase photostability of bleaching fluorophore.
21B xix	Data files not loaded correctly	Mistake in data nomenclature or folder organization	Check that correct input file identifiers are selected. Remove all unnecessary files from parent folders.
	Spot positions do not match images	Wrong format of SPT output; wrong calibration of image files	Re-calibrate images in Fiji. Check that TrackMate output contains spot coordinate columns (X,Y,Z) incl. column headers.
21B xxi	Co-localization rate too low	Channels not aligned properly; bad data quality	Manually re-align channels (see step 21B vii). Repeat image acquisition using improved laser power/gain for limiting channel.
21B xxiii	Corrected number of channel 2 orphans ≠ 0	SNRs in control and reporter image series do not match -	Inspect raw data and choose matching-quality cells. Repeat

		potentially because image acquisition settings do not match	experiment at identical conditions. Re-sort cell lines to achieve similar fluorescence levels.

Timing

Step 1, Design of reporter mRNA: 1-5d
Step 2, Expression of reporter mRNA
Option A, transient transfection: 2d
Option B, stable cell line generation: 28d
Step 3, Preparing samples for northern blotting and qPCR: 2-3d
Steps 4-17, RNA extraction: 2h
Step 18, Detection of RNA degradation using northern blotting and qPCR
Option A, Northern blotting: 2-3d
Steps i-viii, gel electrophoresis: 5h
Steps ix-xxiii, capillary transfer: 5-24h
Steps xxiv-xxxvi, hybridization: 16h – 24h
Step xxxv, quantification: 30min – 1h
Option B, qPCR: 2d
Steps 19-20, Preparing single-molecule imaging experiments: 2-30d
Step 21, Detection of RNA degradation using single molecule imaging
Option A, smFISH: 5d
Steps i-xii, sample preparation: 4d
Steps xiii-xiv, image acquisition: 3h
Steps xv-xxii, spot detection: 3h
Steps xxiii-xxxii, spot co-localization and data analysis: 1h
Option B, Live cell imaging: 3-4d
Steps i-iii, sample preparation: 2d
Steps iv-v, image acquisition: 2h
Steps vi-xiv, single-particle tracking: 1d
Steps xv-xxv, Identification of decay intermediates via track co-localization: 1h

Anticipated Results

This manuscript provides a range of procedures that apply xrRNA-containing reporters for detection and monitoring of mRNA decay. Figure 4 shows the northern blot-based detection of a reporter that is degraded by NMD. In Figure 4b the cells have been transiently transfected with the reporter plasmid as well as a control plasmid (lacZ). Equal signal of the co-transfected control across all samples indicates a consistent transfection efficiency. The signal of the full-length reporter is reduced when a premature stop codon (PTC) is present in the reporter because of its degradation via the NMD pathway. Meanwhile, the faster migrating fragment (xrFrag) corresponds to the XRN1-resistant part of the reporter comprising the xrRNA and the probe binding sites. Note that in the WT control samples this fragment is also visible due to low-level turnover of the reporter. In the PTC samples the xrFrag is enriched, suggesting that 5'-3' exonucleolytic decay is involved in the degradation of the transcript. In Figure 4c the same reporter as in 3b is stably expressed in a HeLa Flp-In TRex cell line. Here, the levels of the endogenous 7SL RNA are detected as a control. In contrast to Figure 4b, the overall signal of the wild type reporter is 3.5% that of the transiently transfected reporter. However,

degradation of the reporter and accumulation of the decay fragment are much more pronounced compared to the transiently transfected cells.

In Figure 4 the same samples as in Figure 4 are analyzed by qPCR which overall leads to similar results. In contrast to the detection by northern blotting, the values are a bit lower. Quantifying the northern blot could possibly cause a slight overestimation of the amounts due to background. We still recommend performing a northern blot because accurate splicing and behavior of the reporter mRNA can be assured this way.

Raw and processed imaging data as well as spot detection and co-localization analysis results obtained from smFISH experiments in cell lines stably expressing reporter and co-localization control transcripts are provided on our lab-website (https://data.fmi.ch/PublicationSupplementRepo/?group=gchao). Representative smFISH images are further shown in Figure 6a. Raw and processed imaging data as well as tracking and track co-localization analysis results obtained from single-molecule live cell experiments in cell lines stably expressing reporter and co-localization control transcripts are provided on the Chao lab-website (https://data.fmi.ch/PublicationSupplementRepo/?group=gchao). Representative live cell images are shown in Figure 6b. For both experimental setups, data can be considered good quality if they yield high detection efficiencies ($e_{C1,C2} \approx 0.9$ for the data provided) since in a control experiment the majority of all particles should be co-localizing and only very few orphans should be detected. To assess whether the quality of control and experimental data is comparable (so that detection efficiencies obtained for control data can indeed be extrapolated to experimental data) inspect the corrected number of channel 2 orphans (provided in each "Results.xlsx" output file in Step 21B xxiv). For the experimental set-up described here, this number corresponds to the corrected number of mRNAs that are labeled with NLS-PCP-GFP only. Since such transcripts should not exist for the described decay reporter, it allows assessment of the quality of the co-localization control data and should be close to 0.

Finally, when examining the overall output (the fraction of intact (co-localizing) transcripts) correlate this value to the half-life (e.g. determined via an alternative method) of the reporter transcript to evaluate the result. For the training data sets, which were acquired 2h after induction, the determined fraction of intact transcripts around 0.5 is in good agreement with the half-life time of the reporter transcript (1.63 ± 0.24 h).

Acknowledgements

This work was funded by the Novartis Research Foundation (J.A.C.), by the Institutional Strategy of the University of Cologne within the German Excellence Initiative (V.B.), a Swiss National Science Foundation (SNF) grant 31003A_156477 (J.A.C.), a grant (GE 2014/4-1) from the Deutsche Forschungsgemeinschaft (N.H.G), the SNF-NCCR RNA & Disease (J.A.C.) and an SNF Marie Heim-Vögtlin fellowship (F.V). N.H.G. acknowledges support by a Heisenberg fellowship (GE 2014/5-1 and GE 2014/7-1). We thank K. Schönig (CIMH) for the parental HeLa 11ht cell line, Elena Dobrikova and Matthias Gromeier for establishing and Matthias Hentze for sharing the HeLa Flp-In T-REx cell line, L. Lavis (Janelia Farm) for providing Halo dyes and L. Gelman and S. Bourke (FMI) for microscopy support as well as H. Kohler (FMI) for cell sorting.

Author Contributions

J.V.G. and V.B. developed northern blotting and qPCR protocols. F.V. and I.H. developed imaging protocols. F.V. developed image processing pipelines with help from J.E. (KNIME). F.V., J.V.G., V.B., I.H., J.A.C and N.H.G. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Data availability statement

Example datasets and analysis software are available from the Chao lab website (https://data.fmi.ch/PublicationSupplementRepo/?group=gchao).

Software 1

Software 1 provides the custom-made KNIME analysis pipeline (termed “Segmentation, Spot Colocalization and Detection Efficiency Correction.knwf”) for analysis of smFISH data.

Software 2

Software 2 provides the custom-made KNIME analysis pipeline (termed “Track-based colocalization analysis (with recursive track pairing and particle number correction)”) for analysis of live cell data.

Data 1

This material provides training data to practice smFISH image analysis, spot detection and co-localization as well as detection efficiency correction (using the KNIME pipeline provided as).

Data 2

This material provides training data to practice live cell data image analysis, tracking and track co-localization as well as detection efficiency correction (using the KNIME pipeline provided as).

References

1 Dolken, L. *et al.* High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959-1972, doi:10.1261/rna.1136108 (2008).

2 Elkon, R., Zlotorynski, E., Zeller, K. I. & Agami, R. Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics* **11**, 259, doi:10.1186/1471-2164-11-259 (2010).

3 Hao, S. & Baltimore, D. The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat Immunol* **10**, 281-288, doi:10.1038/ni.1699 (2009).

4 Schott, J. & Stoecklin, G. Networks controlling mRNA decay in the immune system. *Wiley Interdiscip Rev RNA* **1**, 432-456, doi:10.1002/wrna.13 (2010).

5 Dasgupta, T. & Ladd, A. N. The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA-binding proteins. *Wiley Interdiscip Rev RNA* **3**, 104-121, doi:10.1002/wrna.107 (2012).

6 Marzluff, W. F., Wagner, E. J. & Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9**, 843-854, doi:10.1038/nrg2438 (2008).

7 Schoenberg, D. R. & Maquat, L. E. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* **13**, 246-259, doi:10.1038/nrg3160 (2012).

8 Chen, C. Y. & Shyu, A. B. Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* **2**, 167-183, doi:10.1002/wrna.40 (2011).

9 Dunckley, T. & Parker, R. The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. *EMBO J* **18**, 5411-5422, doi:10.1093/emboj/18.19.5411 (1999).

10 van Dijk, E. *et al.* Human Dcp2: a catalytically active mRNA decapping enzyme located in specific cytoplasmic structures. *EMBO J* **21**, 6915-6924 (2002).

11 Chang, C. T., Bercovich, N., Loh, B., Jonas, S. & Izaurralde, E. The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. *Nucleic Acids Res* **42**, 5217-5233, doi:10.1093/nar/gku129 (2014).

12 Chapman, E. G. *et al.* The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science* **344**, 307-310, doi:10.1126/science.1250897 (2014).

13 Boehm, V., Gerbracht, J. V., Marx, M. C. & Gehring, N. H. Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences. *Nat Commun* **7**, 13691, doi:10.1038/ncomms13691 (2016).

14 Gerbracht, J. V., Boehm, V. & Gehring, N. H. Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells. *Sci Rep* **7**, 10616, doi:10.1038/s41598-017-10847-4 (2017).

15 Horvathova, I. *et al.* The Dynamics of mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells. *Mol Cell* **68**, 615-625 e619, doi:10.1016/j.molcel.2017.09.030 (2017).

16 Wang, C., Han, B., Zhou, R. & Zhuang, X. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. *Cell* **165**, 990-1001, doi:10.1016/j.cell.2016.04.040 (2016).

17 Yan, X., Hoek, T. A., Vale, R. D. & Tanenbaum, M. E. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165**, 976-989, doi:10.1016/j.cell.2016.04.034 (2016).

18 Wu, B., Eliscovich, C., Yoon, Y. J. & Singer, R. H. Translation dynamics of single mRNAs in live cells and neurons. *Science* **352**, 1430-1435, doi:10.1126/science.aaf1084 (2016).

19 Morisaki, T. *et al.* Real-time quantification of single RNA translation dynamics in living cells. *Science* **352**, 1425-1429, doi:10.1126/science.aaf0899 (2016).

20 Pichon, X. *et al.* Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. *J Cell Biol* **214**, 769-781, doi:10.1083/jcb.201605024 (2016).
21 Friedel, C. C., Dolken, L., Ruzsics, Z., Koszinowski, U. H. & Zimmer, R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res* **37**, e115, doi:10.1093/nar/gkp542 (2009).
22 Rabani, M. *et al.* Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29**, 436-442, doi:10.1038/nbt.1861 (2011).
23 Braun, J. E. *et al.* A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. *Nat Struct Mol Biol* **19**, 1324-1331, doi:10.1038/nsmb.2413 (2012).
24 Muhlrad, D., Decker, C. J. & Parker, R. Deadenylation of the unstable mRNA encoded by the yeast MFA2 gene leads to decapping followed by 5'-->3' digestion of the transcript. *Genes Dev* **8**, 855-866 (1994).
25 Matthews, J. C., Hori, K. & Cormier, M. J. Substrate and substrate analogue binding properties of Renilla luciferase. *Biochemistry* **16**, 5217-5220 (1977).
26 Kieft, J. S., Rabe, J. L. & Chapman, E. G. New hypotheses derived from the structure of a flaviviral Xrn1-resistant RNA: Conservation, folding, and host adaptation. *RNA Biol* **12**, 1169-1177, doi:10.1080/15476286.2015.1094599 (2015).
27 Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. & Singer, R. H. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nat Methods* **10**, 119-121, doi:10.1038/nmeth.2305 (2013).
28 Halstead, J. M. *et al.* TRICK: A Single-Molecule Method for Imaging the First Round of Translation in Living Cells and Animals. *Methods Enzymol* **572**, 123-157, doi:10.1016/bs.mie.2016.02.027 (2016).
29 Weidenfeld, I. *et al. Nucleic Acids Research* **37**, e50, doi:10.1093/nar/gkp108 (2009).
30 Grimm, J. B. *et al.* A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods* **12**, 244-250, 243 p following 250, doi:10.1038/nmeth.3256 (2015).
31 Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682, doi:10.1038/nmeth.2019 (2012).
32 Tinevez, J. Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80-90, doi:10.1016/j.ymeth.2016.09.016 (2017).
33 Lionnet, T. *et al.* A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat Methods* **8**, 165-170, doi:10.1038/nmeth.1551 (2011).
34 Dietz, C. & Berthold, M. R. KNIME for Open-Source Bioimage Analysis: A Tutorial. *Adv Anat Embryol Cell Biol* **219**, 179-197, doi:10.1007/978-3-319-28549-8_7 (2016).
35 Trcek, T., Sato, H., Singer, R. H. & Maquat, L. E. Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes & Development* **27**, 551, doi:10.1101/gad.209635.112 (2013).
36 Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-1108 (2008).
37 Halstead, J. M. *et al.* Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. *Science* **347**, 1367-1671, doi:10.1126/science.aaa3380 (2015).
38 Voigt, F., Eglinger, J. & Chao, J. A. Detection of the First Round of Translation: The TRICK Assay. *Methods Mol Biol* **1649**, 373-384, doi:10.1007/978-1-4939-7213-5_25 (2018).

Figure legends

Figure 1: Graphical overview of the experimental workflow

(a) Schematic representation of the xrRNA reporter assay: Reporter transcripts contain xrRNA sequences, which block XRN1-mediated decay. Thus, decay intermediates are stabilized and can be specifically detected along with intact transcripts using the techniques outlined in panel b and c. (b) Overview of experimental steps necessary for molecular biology analysis of xrRNA reporters after RNA extraction. (c) Overview of experimental steps necessary for single-molecule imaging-based detection of xrRNA reporter transcripts.

Figure 2: Reporter design

(a) Schematic overview of the reporter construct design as described in the text. Transcripts must encode a reporter gene open reading frame (ORF) to assay mRNA translation and XRN1-resistant RNA (xrRNA) fragments downstream of the ORF. PP7 and MS2 stem loop cassettes are required for single-molecule imaging and need to be positioned upstream and downstream of the xrRNA sequences. Alternatively, inserting specific probe-binding regions (e.g. part of the β-globin gene) downstream of the xrRNA fragment allows for detection of the stabilized decay intermediate via northern blot analysis (e.g. using a β-globin-specific probe). (b) Reporter constructs can be transiently transfected or stably integrated into the target cell line. (c) Reporter transcript expression leads to production of intact transcripts that can be detected as slow migrating bands via northern blot or as dual-labeled spots via imaging. (d) Processing by XRN1 degrades reporter transcripts. Stalling of XRN1 at xrRNA sites leads to stabilization of decay intermediates that can be detected as faster migrating bands via northern blot or as single-labeled spots via imaging.

Figure 3: Denaturing gel electrophoresis of RNA

Visual inspection of the RNA samples (shown in Figure 3b) after the gel run allows the assessment of successful electrophoresis, equal sample loading and RNA integrity. In an intact RNA sample, the ratio of 28S and 18S rRNA should be approximately 2:1 (can be 1:1 for certain tissues and cell lines). No apparent smearing in the lanes should be visible before the transfer, the gel is cut according to the expected size of the transcript of interest and the control RNA. For the detection of RNA fragments ranging from 0.3 – 7 kb, cut at the indicated positions. PTC = premature stop codon.

Figure 4: Northern blot xrRNA analysis of transcripts degraded by NMD

(a) Schematic depiction of the reporter mRNA. The reporter consists of the β-globin coding sequence (exons shown as orange boxes). The introduction of a premature stop codon (PTC) in addition to the normal stop codon (stop) targets the transcript for NMD. Downstream, the Murray valley encephalitis (MVE) xrRNA is placed followed by the northern probe binding site. Stalling of XRN1 at the xrRNA gives rise to a decay intermediate fragment that can be detected on a northern blot (xrFrag). (b) Northern blot of triplicate RNA samples derived from HeLa cells that have been either transiently transfected with the vector pcDNA5/FRT/TO-globinWT-xrRNA4H or pcDNA5/FRT/TO-globinPTC-xrRNA4H (Addgene IDs 108375 and 108376). As control for the transient transfection, a lacZ expressing plasmid was co-transfected and used for normalization. Below the reporter, a faster migrating fragment (xrFrag) is visible that

results from stalling of XRN1 at the xrRNA. (c) The signal intensities of the reporter and xrFrag shown in (b) were quantified and normalized to the wild type controls. The values are plotted as individual data points and the horizontal bar represents the mean value of each replicate (n=3). (d) Northern blot analysis of RNA samples derived from stable HeLa Flp-In T-REx cells that express the reporter from the integrated genomic locus. 7SL serves as an endogenous control. (e) The signal intensities of the reporter and xrFrag shown in (d) were quantified and normalized to the wild type controls. The values are plotted as individual data points and the horizontal bar represents the mean value of each replicate (n=3).

Figure 5: qPCR xrRNA analysis of transcripts degraded by NMD

(a) The analysis was performed for the reporter shown in Figure 3a. One set of primers amplifies the globin reporter. The primers are placed in different exons in order to avoid amplification of genomic DNA. The XRN1-resistant fragment is amplified by primers that bind in the 5' linker and in a unique sequence in the xrRNA element. (b) For samples from transient transfections, the fold change of reporter and xrFrag were calculated by first normalizing to the internal co-transfected mVenus control and then normalizing to the respective wild type control. The values are plotted as individual data points and the horizontal bar represents the mean value of each replicate (n=2). (c) For samples from stable cell lines the fold changes of reporter and xrFrag were calculated by first normalizing to the endogenous GAPDH control and then normalizing to the wild type control. The values are plotted as individual data points and the horizontal bar represents the mean value of each replicate (n=2).

Figure 6: Representative smFISH and live cell imaging data and analysis steps

(a) Representative smFISH image of a HeLa cell stably expressing NLS-MCP-Halo, NLS-PCP-GFP and an xrRNA reporter transcript hybridized with probes against the MS2-containing 3'UTR (Ch1, magenta) as well as ORF (Ch2, green). The nucleus was stained using DAPI. Cells were induced by addition of doxycycline for 2 h and fixed 2 h after induction was stopped. Dashed yellow lines illustrate the region of interest (ROI) for spot detection. Magenta and green circles mark detected spots in Ch1 and Ch2. Co-localizing spots represent intact reporter transcripts. Magenta-only (orphan) spots are stabilized degradation intermediates. Scale bar = 5 µm. (b) Representative live cell image of a HeLa cell stably expressing NLS-MCP-Halo, NLS-PCP-GFP and an xrRNA reporter transcript approximately 2 h after induction of expression. Dashed yellow lines illustrate the region of interest (ROI) for spot detection. Magenta and green circles mark detected spots in Ch1 and Ch2 that were joined into tracks (magenta and green lines) when combining spot coordinates from multiple frames. Co-localizing tracks are intact reporter transcripts. Orphan tracks are stabilized degradation intermediates. Scale bar = 5 µm.

Boxes

Box 1: Detection efficiencies and particle number correction

For a co-localization control experiment, the detection efficiency ($e_{C1,C2}$) (as originally derived by Horvathova et al. (2017)) can be defined as follows:

$$e_{C1} = \frac{n_{C1} + n_{Coloc}}{N}$$

And:

$$e_{C2} = \frac{n_{C2} + n_{Coloc}}{N}$$

with the total number of particles $N = n_{C1} + n_{C2} + n_{Coloc}$. Using the detection efficiencies defined above, the corrected fraction of co-localizing particles (f_{Coloc}) can be determined by application of the true underlying particle numbers ($m_{C1,C2,Coloc}$) as follows:

$$m_{C1} = \frac{n_{C1}}{e_{C1}} - m_{Coloc}(1 - e_{C2})$$

And:

$$m_{C2} = \frac{n_{C2}}{e_{C2}} - m_{Coloc}(1 - e_{C1})$$

For determining the true underlying particle numbers:

$$m_{Coloc} = \frac{n_{Coloc}}{e_{C1}e_{C2}}$$

For determining the corrected fraction of co-localizing particles:

$$f_{Coloc} = \frac{m_{Coloc}}{m_{C1} + m_{C2} + m_{Coloc}}$$

Supplementary Material

Supplementary Procedure 1. Radioactive norther probe labeling.

This supplementary procedure provides a protocol for radioactive northern probe labeling.

Supplementary Procedure 2. Non-radioactive northern probe labeling

This supplementary procedure provides a protocol for non-radioactive northern probe labeling using biotin.

Supplementary Procedure 3. Stellaris FISH probe design.

This supplementary material provides instructions that aim to help with Stellaris FISH probe design.

Supplementary Table 1. Plasmid information

This table provides an overview of the plasmids used throughout the procedure.

Supplementary Table 2. Oligonucleotide sequences

This table provides an overview of the oligonucleotide sequences applied throughout the protocol.

Figure 1

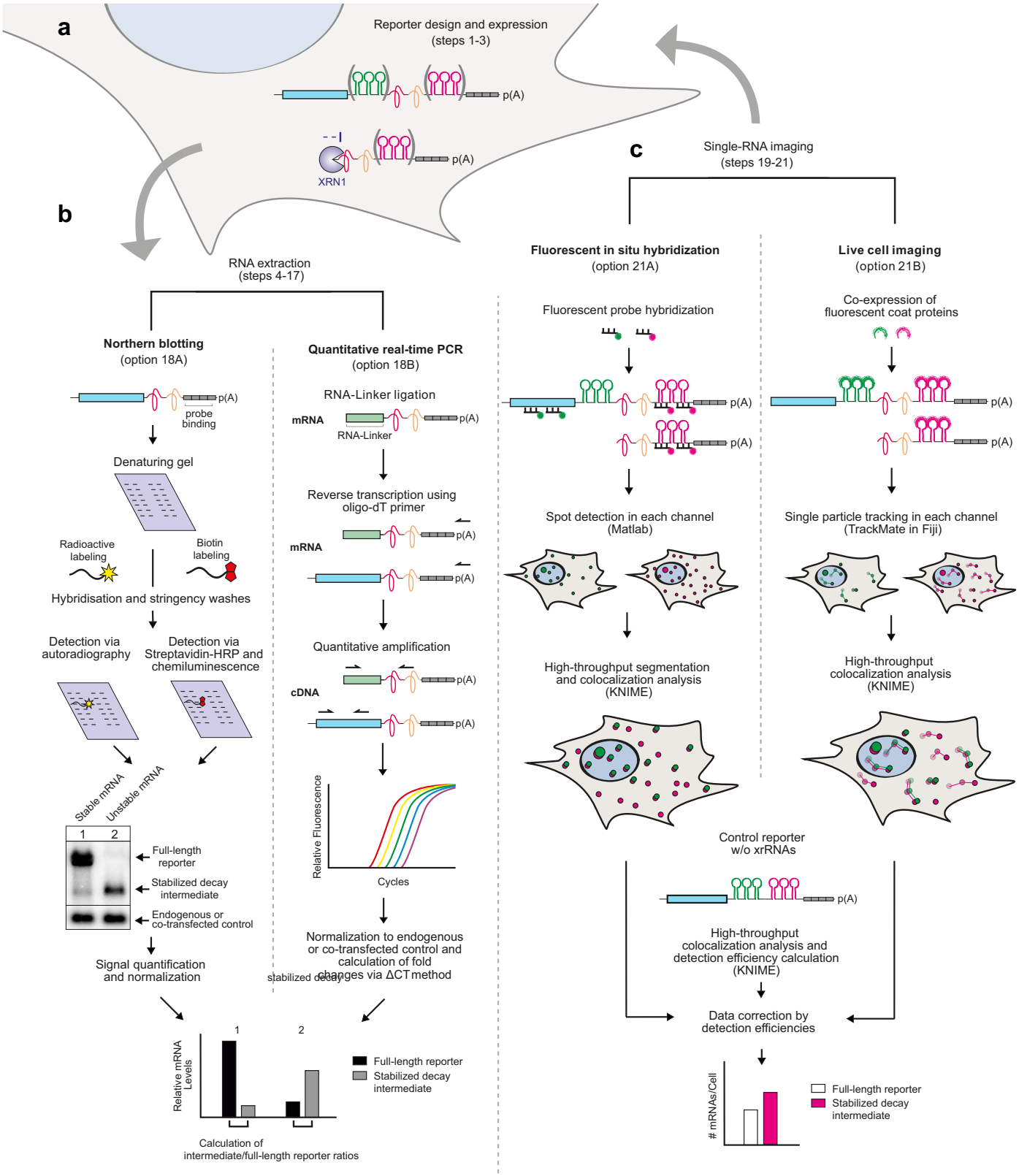

a
Reporter design and expression
(steps 1-3)
p(A)
XRN1
b
RNA extraction
(steps 4-17)
Northern blotting
(option 18A)
probe binding
Denaturing gel
Radioactive labeling
Biotin labeling
Hybridisation and stringency washes
Detection via autoradiography
Detection via Streptavidin-HRP and chemiluminescence
Stable mRNA
Unstable mRNA
Full-length reporter
Stabilized decay intermediate
Endogenous or co-transfected control
Signal quantification and normalization
Quantitative real-time PCR
(option 18B)
RNA-Linker ligation
mRNA
RNA-Linker
Reverse transcription using oligo-dT primer
Quantitative amplification
cDNA
Relative Fluorescence
Cycles
Normalization to endogenous or co-transfected control and calculation of fold changes via ΔCT method
stabilized decay
Relative mRNA Levels
Calculation of intermediate/full-length reporter ratios
c
Single-RNA imaging
(steps 19-21)
Fluorescent in situ hybridization
(option 21A)
Fluorescent probe hybridization
Spot detection in each channel
(Matlab)
High-throughput segmentation and colocalization analysis
(KNIME)
Live cell imaging
(option 21B)
Co-expression of fluorescent coat proteins
Single particle tracking in each channel
(TrackMate in Fiji)
High-throughput colocalization analysis
(KNIME)
Control reporter
w/o xrRNAs
High-throughput colocalization analysis and detection efficiency calculation
(KNIME)
Data correction by detection efficiencies
mRNAs/Cell
Full-length reporter
Stabilized decay intermediate

Figure 2

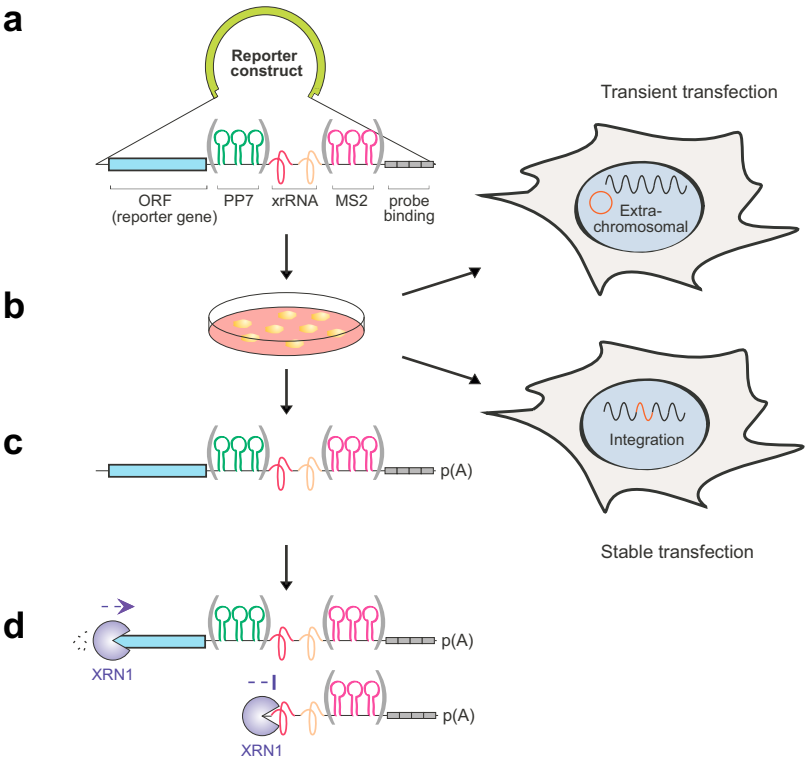

a
Reporter construct
ORF (reporter gene)
PP7
xrRNA
MS2
probe binding
b
c
p(A)
d
XRN1
p(A)
XRN1
p(A)
Transient transfection
Extra-chromosomal
Integration
Stable transfection

Figure 3

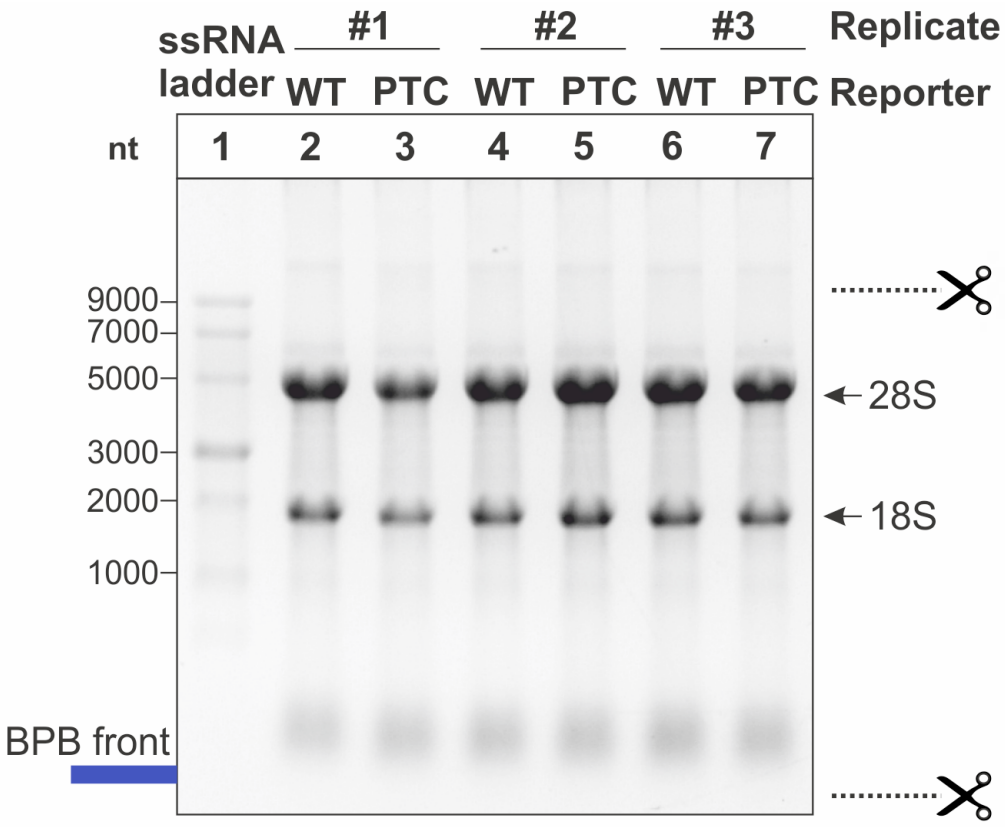

ssRNA ladder
#1
#2
#3
Replicate
WT PTC WT PTC WT PTC
Reporter
nt
1 2 3 4 5 6 7
9000
7000
5000
3000
2000
1000
28S
18S
BPB front

Figure 4

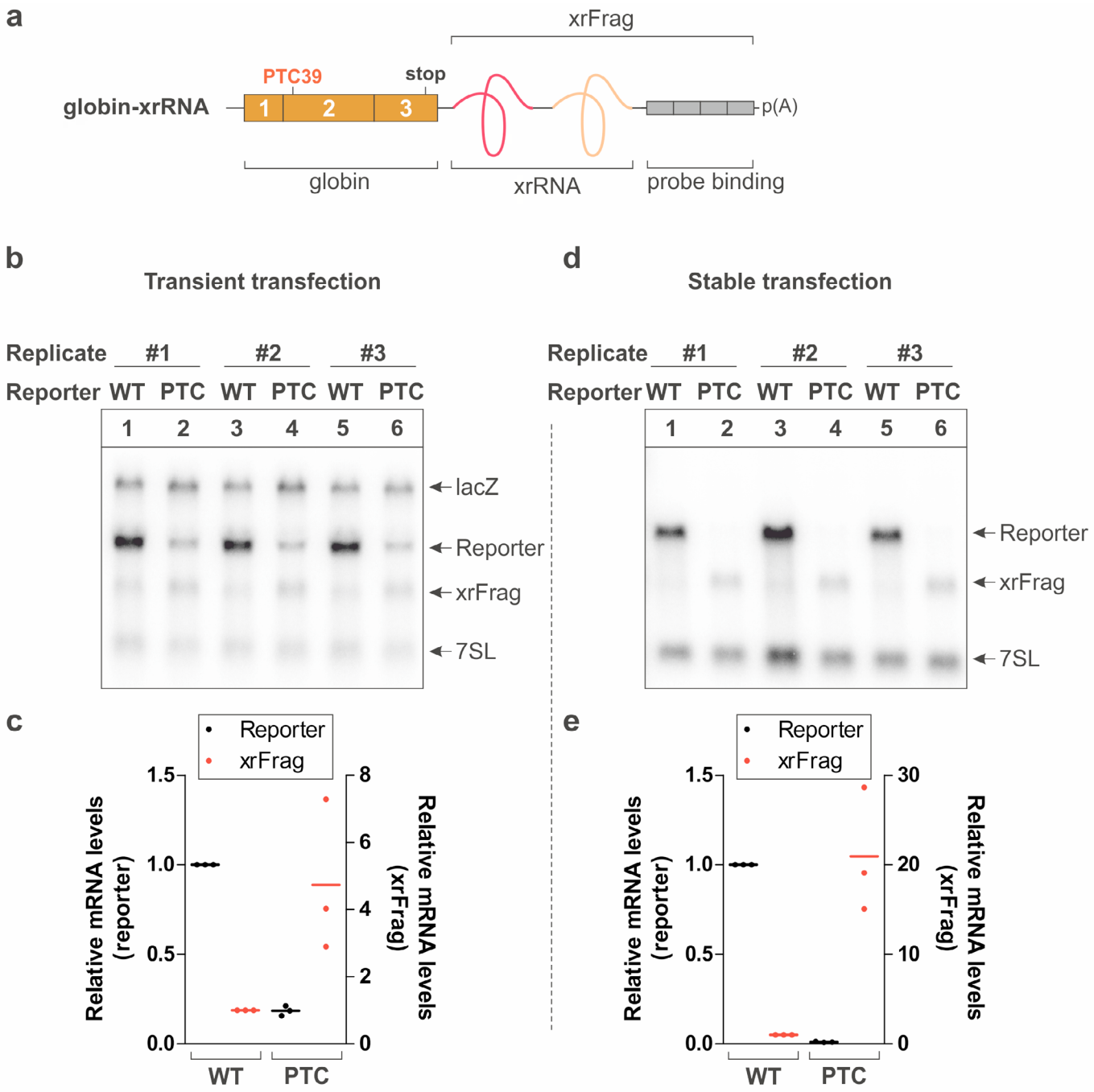

a
xrFrag
PTC39
stop
globin-xrRNA
1
2
3
p(A)
globin
xrRNA
probe binding
b
Transient transfection
Replicate #1 #2 #3
Reporter WT PTC WT PTC WT PTC
1 2 3 4 5 6
lacZ
Reporter
xrFrag
7SL
d
Stable transfection
Replicate #1 #2 #3
Reporter WT PTC WT PTC WT PTC
1 2 3 4 5 6
Reporter
xrFrag
7SL
c
Reporter
xrFrag
Relative mRNA levels (reporter)
Relative mRNA levels (xrFrag)
1.5
1.0
0.5
0.0
8
6
4
2
0
WT
PTC
e
Reporter
xrFrag
Relative mRNA levels (reporter)
Relative mRNA levels (xrFrag)
1.5
1.0
0.5
0.0
30
20
10
0
WT
PTC

Figure 5

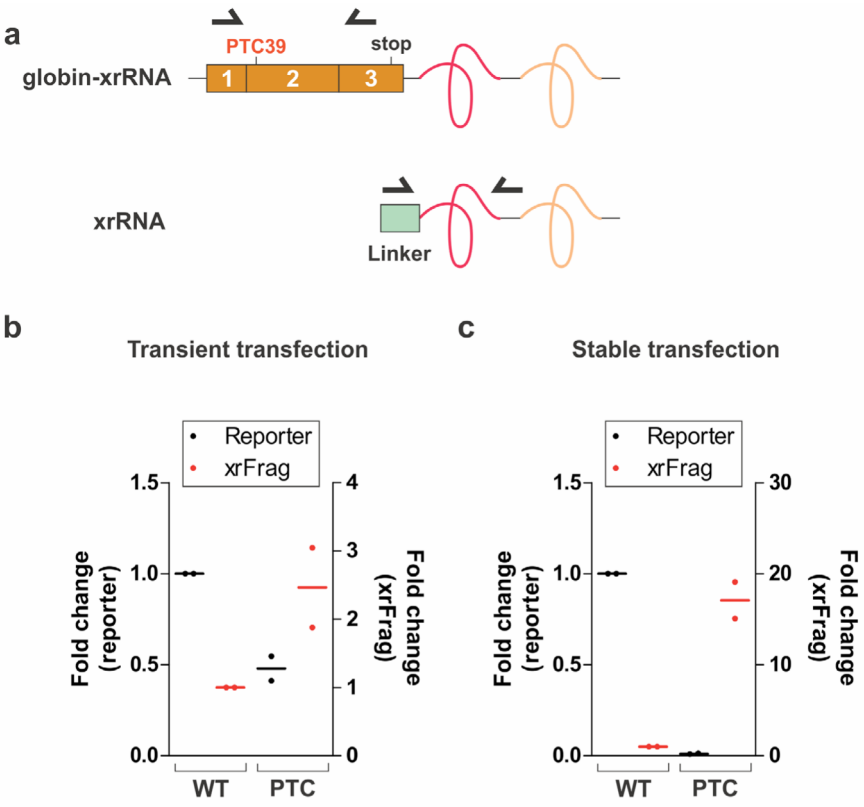
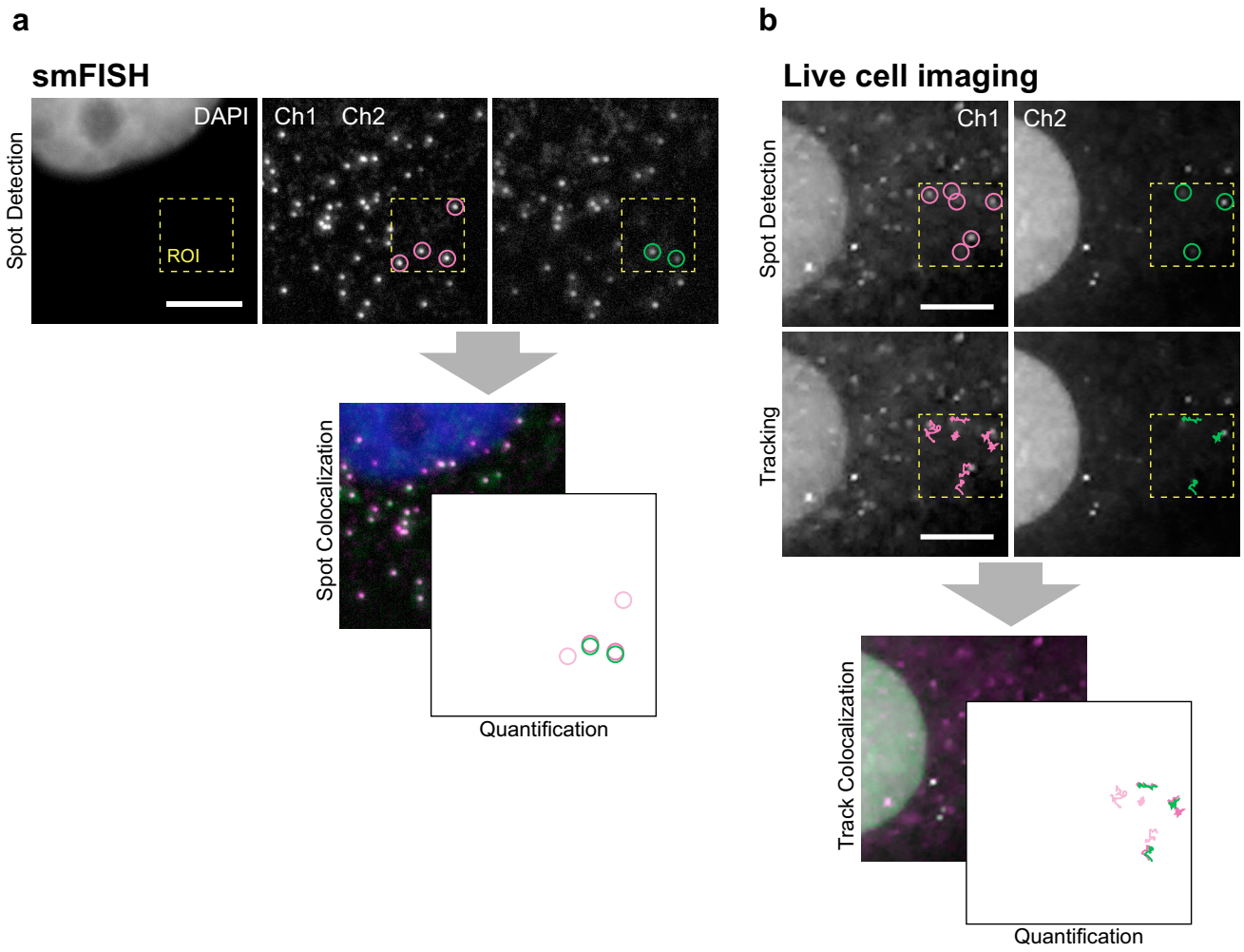

a
PTC39
stop
globin-xrRNA
1
2
3
xrRNA
Linker
b
Transient transfection
Reporter
xrFrag
Fold change (reporter)
Fold change (xrFrag)
1.5
1.0
0.5
0.0
4
3
2
1
0
WT
PTC
c
Stable transfection
Reporter
xrFrag
Fold change (reporter)
Fold change (xrFrag)
1.5
1.0
0.5
0.0
30
20
10
0
WT
PTC

Figure 6


a
smFISH
Spot Detection
DAPI
ROI
Ch1
Ch2
Spot Colocalization
Quantification
b
Live cell imaging
Spot Detection
Ch1
Ch2
Tracking
Track Colocalization
Quantification

Supplementary Procedure 1. Radioactive northern probe labeling

The probe binding site of the β-globin reporter contains four repeats of a 100 nt sequence from the β-globin gene. These binding sites can be placed in any reporter of choice to enable detection by the radioactive β-globin northern blot probe (e.g. Addgene #120390).

1. For the *in vitro* transcription of the radioactive β-globin northern blot probe, mix the following reagents in a nuclease free 1.5 ml microcentrifuge tube and incubate for 40 min at 40 °C on a heat block.

Reagents	Volume (µl) per reaction
10x transcription buffer	2
Ultra pure water	4
0.1 M DTT	0.5
10 mM ATP	1.2
10 mM CTP	1.2
10 mM UTP	1.2
RNasin (40 U/µl)	0.7
SP6 RNA polymerase	1.2
BamHI linearised plasmid DNA (approx. 50 ng/µl) (e.g. Addgene #116927)	3
α-^{32}P-GTP (10 µC/µl)	5
Total reaction volume	20

2. Add 1.5 µl of RNase-free DNAse I (1 U/µl), mix well and incubate for 20 min at 37 °C on a heat block.
3. Add 28.5 µl of ultra pure water, mix well.
4. Purify the transcribed probe using the mini Quick Spin RNA Columns following the manufacturer's instructions.
5. Use a Geiger counter to confirm the successful transcription and purification of the probe. Roughly equal radioactive signal in the column and flow through indicate good transcription efficiency. Alternatively, check the transcription by denaturing PAGE.
6. Dilute the probe in 20-25 ml of Church buffer and use immediately for hybridization or store at -20 °C for up to one week.
7. For the radioactive labeling of the 7SL control northern blot probe, mix the following reagents in a nuclease free 1.5 ml microcentrifuge tube and incubate for 60 min at 37 °C on a heat block.

Reagents	Volume (µl) per reaction
10 mM 7SL DNA oligo	1
10x PNK Buffer A (forward reaction)	2
Ultra pure water	11
γ-^{32}P-ATP (10 µC/µl)	5
PNK enzyme	1
Total reaction volume	20

8. Add 30 µl of ultra pure water, mix well.
9. Purify the transcribed probe using the mini Quick Spin RNA Columns following the manufacturer's instructions.
10. Use a Geiger counter to confirm the successful transcription and purification of the probe.
11. Dilute the probe in 20-25 ml of Church buffer and use immediately for hybridization or store at -20 °C up to one week.

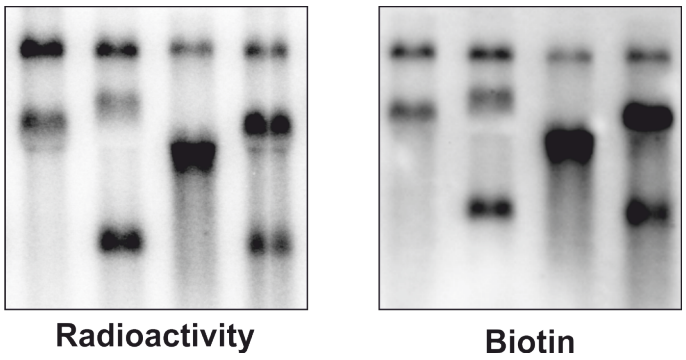
Supplementary Procedure 2. Non-radioactive northern probe labeling

1. Perform *in vitro* transcription to generate a biotin-labeled β-globin northern blot probe by mixing the following reagents in a nuclease free 1.5 ml microcentrifuge tube and incubate for 40 min at 40 °C on a heat block.

Reagents	Volume (µl) per reaction
10x transcription buffer	2
Ultra pure water	7.8
0.1 M DTT	0.5
10 mM ATP	1.2
10 mM CTP	1.2
10 mM GTP	1.2
10 mM Biotin-11-UTP	1.2
RNasin (40 U/µl)	0.7
SP6 RNA polymerase	1.2
BamHI linearised plasmid DNA (approx. 50 ng/µl)	3
Total reaction volume	20

2. Add 1.5 µl of RNase-free DNAse I (1 U/µl), mix well and incubate for 20 min at 37 °C on a heat block.
3. Add 28.5 µl ultra-pure H_2O and mix well.
4. Purify the transcribed probe using the mini Quick Spin RNA Columns following the manufacturer's instructions.
5. Check for successful *in vitro* transcription by running 5 µl of the reaction mix on a denaturing PAGE.
6. Dilute the probe in 20-25 ml of Church buffer and use immediately for hybridization or store at -20 °C for up to one year.
7. To make a probe for the detection of 7SL RNA, dilute a 5' biotinylated DNA oligo in Church buffer to a final concentration of 50 nM. E.g. dilute 12.5 µl of a 100 µM DNA oligo stock solution in 25 ml Church buffer.
8. Hybridize the blot overnight at 40°C in a hybridization oven while rotating
9. Store the hybridization solution for later use at -20 °C (Signal could be weaker when re-used)
10. Wash the membrane for 15 min with 20-30 ml of northern wash buffer 1 while rotating at 40 °C in a hybridization oven.
11. Discard the washing solution and repeat step 10.
12. Discard the washing solution and wash the membrane for 15 min with 20-30 ml of northern wash buffer 2 while rotating at 40 °C in a hybridization oven.
13. Discard the washing solution and repeat step 12.
14. Equilibrate the membrane for 2 min in TBS-T.
15. Block the membrane in 5% BSA TBS-T on a rocking platform for 1h.
16. Incubate the membrane in a 1:10,000 dilution of streptavidin-HRP in 5% BSA TBS-T.
17. Wash the membrane 3 times for 10 min each in TBS-T.
18. Detect the signal by adding ECL to the membrane and visualize in a chemiluminescence imager.

19. Example gel image for comparison: See below, image shows the same samples from cells transiently expressing reporter mRNAs detected by radioactive (left) and biotin (right) labeled probes.



Radioactivity Biotin

Supplementary Procedure 3. Stellaris FISH probe design

- For FISH probe design, use the Biosearch Technologies web portal: https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer

- Design two sets of DNA oligo probes (up to 48 probes (app. 20 nt long) per set, if sequence is long enough use masking level 5) labeled with spectrally distinct fluorophores (e.g. Quasar570 and Quasar670).
- If possible, test both fluorophores with both probe sets to see which combination yields best labeling efficiencies on the microscopic set-up used.
- Each probe set needs to target a region of interest downstream and upstream of the xrRNA fragments so that stabilized RNA intermediates are recognized by only a single probe set while intact mRNAs are recognized by both probe sets.
- Specifically, we designed probes against MS2 repeats and the Renilla luciferase ORF. Probe sequences were published as supplemental data by Horvathova et al. (2017).

Supplementary Table 1. Plasmid information

Addgene ID	Plasmid name	Features
108366	pCI-globin_del5UTR_WT-xrRNA-4H	Wild type beta-globin reporter with MVE xrRNA and 4H probe binding sites (Homo sapiens)
108367	pCI-globin_del5UTR_PTC39-xrRNA-4H	PTC39 beta-globin reporter with MVE xrRNA and 4H probe binding sites (Homo sapiens)
108368	pCI-TPI_WT-xrRNA-4H	Wild type TPI reporter with MVE xrRNA and 4H probe binding sites (Homo sapiens)
108369	pCI-TPI_PTC160-xrRNA-4H	PTC160 TPI reporter with MVE xrRNA and 4H probe binding sites (Homo sapiens)
108370	pCI-TPI_WT-xrRNA-4MS2-4H	Wild type TPI reporter with 4MS2 binding sites downstream of MVE xrRNA and 4H probe binding sites (Homo sapiens)
108371	pCI-TPI_WT-4MS2-xrRNA-4H	Wild type TPI reporter with 4MS2 binding sites upstream of MVE xrRNA and 4H probe binding sites (Homo sapiens)
108372	pCI-TPI_WT-RAB7A_3UTR-xrRNA-4H	Wild type TPI reporter with partial RAB7A control 3' UTR, MVE xrRNA and 4H probe binding sites (Homo sapiens)
108373	pCI-TPI_WT-IL6_3UTR-xrRNA-4H	Wild type TPI reporter with partial IL6 control 3' UTR, MVE xrRNA and 4H probe binding sites (Homo sapiens)
108374	pCI-TPI_WT-TNFalpha_3UTR-xrRNA-4H	Wild type TPI reporter with partial TNFalpha control 3' UTR, MVE xrRNA and 4H probe binding sites (Homo sapiens)
108375	pcDNA5/FRT/TO-globin_del5UTR_WT-xrRNA-4H	Wild type beta-globin reporter with MVE xrRNA and 4H probe binding sites (Homo sapiens)
108376	pcDNA5/FRT/TO-globin_del5UTR_PTC39-xrRNA-4H	PTC39 beta-globin reporter with MVE xrRNA and 4H probe binding sites (Homo sapiens)
108377	pcDNA5/FRT/TO-TPI_WT-xrRNA-4H	Wild type TPI reporter with MVE xrRNA and 4H probe binding sites (Homo sapiens)
108378	pcDNA5/FRT/TO-TPI_PTC160-xrRNA-4H	PTC160 TPI reporter with MVE xrRNA and 4H probe binding sites (Homo sapiens)
120390	pGEM4Z globin NB Probe	Allows in vitro SP6 Polymerase driven transcription of a beta globin northern blot probe
104095	pSF4 TetCMV intron Renilla-STOP 24xPP7 1xfirefly-siRNA xrRNA12 24xMS2 SV40 CTE polyA	Expresses xrRNA reporter transcripts in mammalian cells
104097	pSF4 TetCMV intron Renilla-STOP 24xPP7 1xfirefly-siRNA 24xMS2 SV40 CTE polyA	Expresses co-localization control transcripts in mammalian cells
104098	phage UbiC NLS HA stdMCP Halo	Lentiviral vector expressing synonymous tandem MS2 coat protein fused to Halo. Used to label mRNAs.
104099	phage UbiC NLS HA stdPCP stdGFP	Lentiviral vector expressing synonymous tandem PP7 coat protein fused to synonymous tandem GFP. Used to label mRNAs.

Supplementary Table 2. Oligonucleotide sequences

Name	Type	Step	Sequence
Human 7SL	DNA	Supplementary Procedure 1, step 7	TGCTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTT
Human 7SL biotinylated	5' biotinylated DNA	Supplementary Procedure 2, step 7	/5BiosG/TGCTCCGTTTCCGACCTGGGCCGGTTCACCCCTCCTT
RNA linker VBnew2	RNA	Step 18	rGrCrUrGrArUrGrGrCrGrArUrGrArArUrGrArNrNrNrNrNrNrArArA
VNN-oligo-dT20	DNA	Step 18	TTTTTTTTTTTTTTTTTTTTVNN
Globin_qPCR_se	DNA	Step 18	AAGGCTCATGGCAAGAAAG
Globin_qPCR_as	DNA	Step 18	ACACCAGCCACCACTTTC
RNA_Linker_VB2_only_se	DNA	Step 18	GCTGATGGCGATGAATGA
MVE-qPCR_as	DNA	Step 18	CCGTGGAGAATCAGCTTTGG

3 DISCUSSION

Post-transcriptional regulation is carried out via multiple RNA species, messenger ribonucleoproteins and complex protein machineries across different cellular departments. In Figure 7 an overview of these processes and the related published results is provided. The publications underlying this cumulative thesis have analyzed and increased the understanding of several major aspects of post-transcriptional gene regulation: 1) the characterization of a factor involved in the processing of ribosomal RNAs, an essential part of the protein biosynthesis machinery (Saez *et al.*, 2020); 2) the splicing-dependent assembly of the EJC in the nucleus (Gerbracht & Gehring, 2018); 3) mechanisms of splicing regulation by the EJC and associated factors (Boehm *et al.*, 2018, Gerbracht *et al.*, 2019); 4) NMD activation by the EJC (Gerbracht *et al.*, 2019); 5) the contribution of degradation pathways to the turnover of unstable mRNAs (Boehm *et al.*, 2016, Gerbracht *et al.*, 2017). Finally, using XRN1-resistant sequences in reporter mRNAs has been established as a powerful tool to monitor XRN1-dependent degradation (Boehm *et al.*, 2016, Gerbracht *et al.*, 2017, Voigt *et al.*, 2019).

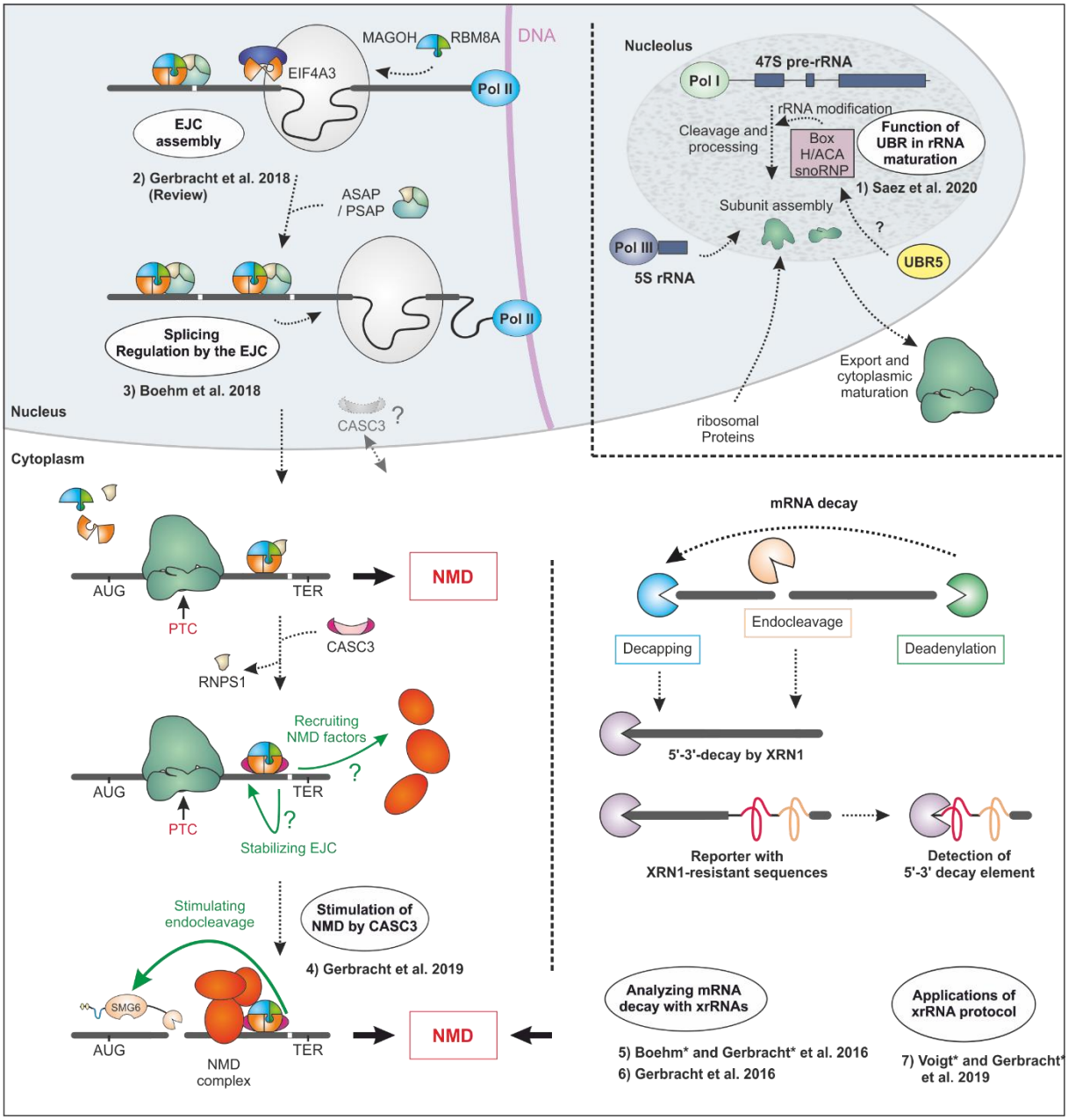


Figure 7: Biological processes underlying the publications of this thesis. Schematic depiction of cellular processes and their subcellular location. The references to publications are in order of their appearance in the results section.

3.1 A link between the E3 ubiquitin ligase UBR5 and ribosomal RNA maturation in embryonic stem cells

Human embryonic stem cells have a distinct molecular makeup to maintain pluripotency. One of the post-transcriptional mechanisms involved is the ubiquitin proteasome system (Saez *et al.*, 2018). Among other ubiquitin proteasome factors, the E3-ubiquiting ligase UBR5 plays a role in preserving the expression of pluripotency factors while inhibiting the expression of

differentiation factors (Buckley *et al.*, 2012). Correspondingly, UBR5 expression decreases with differentiation (Koyuncu *et al.*, 2018).

In Saez *et al.* (2020), an additional layer of UBR5 function in embryonic stem cells is reported. UBR5 directly interacts with the proteins GAR1, NHP2 and NOP10 of the H/ACA snoRNP in an RNA-independent manner. However, UBR5 does not alter the stability of the H/ACA snoRNP proteins, suggesting a role beyond its function as a E3 ubiquiting ligase. Furthermore, UBR5 interacts with the snoRNAs SNORA73A/U17 and SNORA5A which are members of the H/ACA snoRNP. Both SNORA73A and SNORA5A are implicated in the pseudouridylation of sites present in the 18S rRNA (Atzorn *et al.*, 2004, Kiss *et al.*, 2004). UBR5 has been previously shown to interact with proteins that play a role in a great variety of cellular processes such as the cell cycle, transcriptional regulation and the DNA damage response. Not all of these are targeted by UBR5's ligase activity, suggesting that UBR5-mediated ubiquitination is regulated (Shearer *et al.*, 2015).

Tafforeau *et al.* (2013) have performed an extensive screen to identify processing factors involved in rRNA processing. In this study, 625 proteins nucleolus-localizing proteins were knocked down individually and the effects on rRNA intermediates was observed. Using this approach, 286 factors involved in human rRNA processing were identified. UBR5 does not localize to the nucleolus and was therefore not included in this screen. Considering the defects in rRNA maturation that are observed upon UBR5 depletion, it is conceivable that an even larger number of hitherto unidentified factors contributes to rRNA maturation.

So far, the analysis of UBR5's function in rRNA processing has been limited to human and murine embryonic stem cells. Since in these cell types UBR5 is upregulated compared to differentiated cells the question remains whether UBR5 functions similarly in rRNA processing of differentiated cells.

3.2 New insights into the composition and function of the EJC

Since the discovery of the EJC, a great variety of EJC-interacting factors have been identified and their functions explored. However, many open questions remain. This includes the temporal dynamics of EJC assembly and the post-transcriptional EJC-interactome and related functions. Analyzing these questions with state-of-the-art technologies such as high-throughput omics, CRISPR-Cas9 genome editing and cryo-EM has led to a new understanding

of many aspects of EJC biology. In Gerbracht & Gehring (2018), the recently published cryo-EM structures of the human spliceosome are discussed with a focus on interactions between the EJC core components and spliceosomal proteins. In Boehm *et al.* (2018) a transcriptome-wide analysis of EJC-dependent alternative splicing reveals that the EJC together with the splice factor RNPS1 protects splicing fidelity. In Gerbracht *et al.* (2019) CRISPR-Cas9-generated knockout cells and subsequent mass-spectrometry and RNA-sequencing are used to show that CASC3's primary molecular function is not as part of the EJC core as previously believed. Instead, CASC3's main role is as a mediator between the EJC and the NMD machinery, possibly as an NMD activating branch besides UPF3B and/or RNPS1.

3.2.1 Nuclear EJC: Assembly and splicing

Since the publication of the review Gerbracht & Gehring (2018), three additional human spliceosomal structures of intermediates containing the EJC or EJC-related factors have been published (Haselbach *et al.*, 2018, Zhan *et al.*, 2018, Zhang *et al.*, 2018, Fica *et al.*, 2019). In the following they will be discussed to provide an updated view of EJC assembly in connection with splicing (Figure 8). The B^{act} spliceosome is the earliest stage in which CWC22 is present, in agreement with an earlier mass spectrometry-based analysis (Bessonov *et al.*, 2010). CWC22 flanks the 5' exon with its MA3 and MIF4G domains, similarly as in the C* structure. The interactions observed in C* between MA3 and PRPF8 (Prp8 in yeast) on the one hand and MIF4G and EFTUD2 (Snu114) on the other hand are already established in B^{act}. There is a newly identified interaction between CWC22's MA3 domain and RNF113A (Cwc24) that is specific for the B^{act} complex. RNF113A as part of the spliceosome core connects the guanosine residue at the 5' end of the intron, PRPF8, CWC27, SNRNP200 (Brr2), and the U2 snRNP. Together with the U2 snRNP it separates the 5' SS from the catalytic center while preventing a premature nucleophilic attack by the branch point adenosine. During catalytic activation by DHX16 (Prp2), RNF113A and CWC27 are released and are already missing in the structure of the late B^{act} complex (Zhang *et al.*, 2018). This results in a conformational change of PRPF8 that secures exon binding. (Kastner *et al.*, 2019). At the same time, the 5' SS is free to access the catalytic center allowing the branching reaction to occur. The structure of the B* spliceosome is the only stage still missing and it is therefore unclear if and how the MAGOH-RBM8A heterodimer associates with EIF4A3 in the B* spliceosome. In the C, C* and post-catalytic complex, all core EJC components as well as parts of CASC3 are resolved (Zhan *et al.*, 2018, Fica *et al.*, 2019).

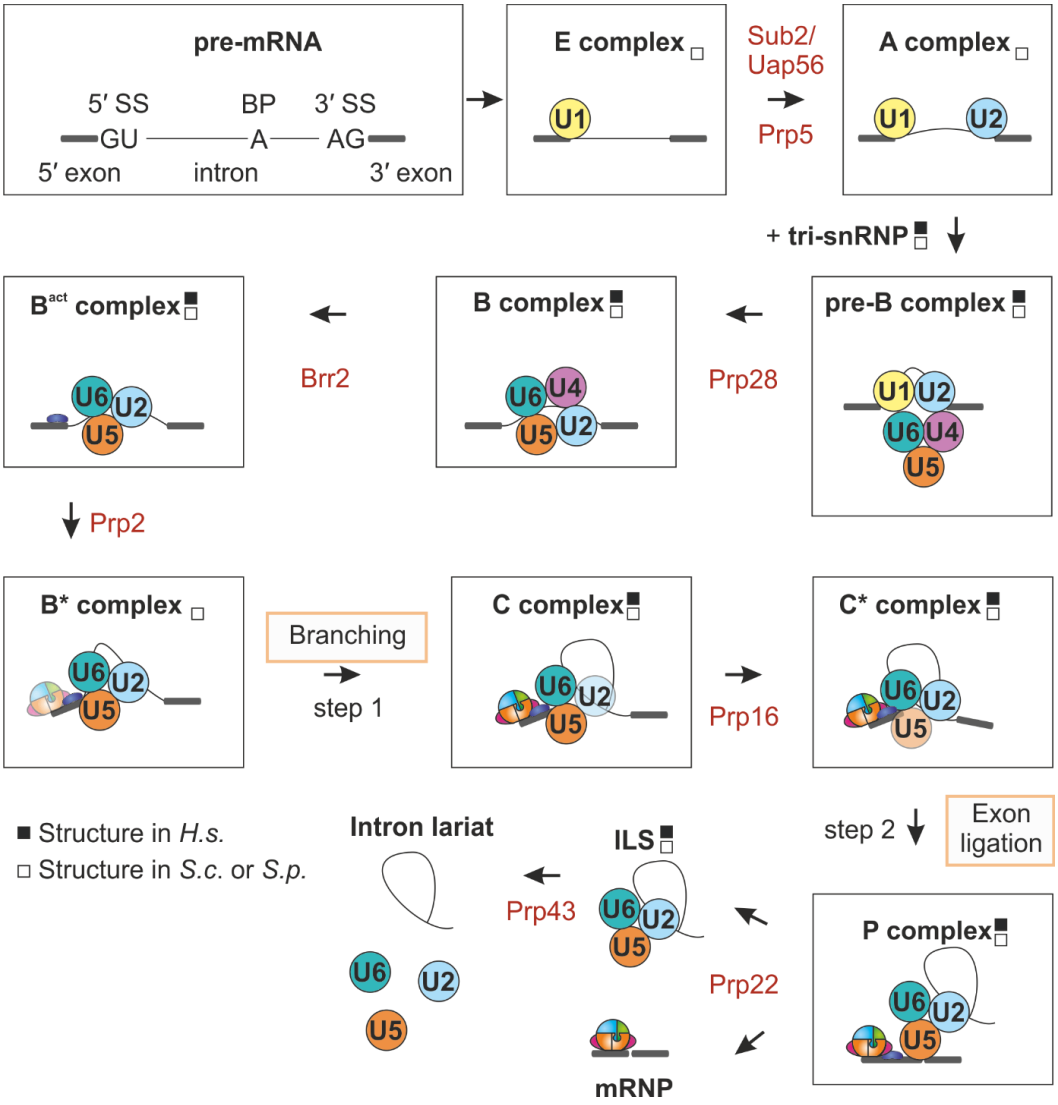


Figure 8: Rearrangements of the spliceosome core during the splicing reaction. *Shown are the intermediates of the splicing reaction with a focus on the changes of snRNP composition and the assembly of the EJC (depicted as in Figure 4). CWC22 is shown as a purple oval shape. In red, the yeast ATP-dependent helicases are specified that cause conformational changes within the spliceosome (conserved in the human spliceosome). The 5′ splice site (5′ SS) is bound by the U1 snRNP. In the A complex, the branch point site (BP) is recognized and bound by the U2 snRNP. The binding of the pre-formed tri-snRNP formed of U4, U5 and U6 results in the pre-B complex. CWC22 joins the spliceosome in the activated B-complex. The catalytically activated B* complex is ready to perform the first step of the splicing reaction: the nucleophilic attack of the 5′ SS on the adenosine at the BP. Since no human B* spliceosomal structure has been published yet, the composition of the EJC at this time point is not known. In the C complex, the EJC is present at the 5′ exon. The catalytically active C* spliceosome is primed to catalyze the second step of the splicing reaction, the exon ligation reaction. The post-splicing complex (P complex) is disassembled by Prp22 (DHX8 in human) which includes the disassociation of CWC22. The exon junction complex stays bound upstream of the exon-exon junction in the mature messenger ribonucleoprotein. The intron lariat structure (ILS) is then disassembled by PRP43 (DHX15 in human). The black and white squares indicate whether the complexes have been structurally resolved in human (*homo sapiens, h.s.*), or yeast (*Saccharomyces cerevisiae, S.c. *or* Saccharomyces pombe, S.p.*).*

Taken together, the structures of the human spliceosome published so far have revealed new insights into the sequential assembly of CWC22 and the EJC and many new interaction sites.

This includes the contacts between CWC22 and the spliceosomal components RNF113A, EFTUD2, PRPF8 and SLU7. To further investigate the functions of these interactions, CWC22 mutated at the individual contact sites could be expressed in CWC22-depleted cells and the effect on splicing of endogenous targets observed by performing RT-PCT. These mutants could also be tested for splicing efficiency of a labelled transcript in an *in vitro* splicing assay (Steckelberg & Gehring, 2014). A previously unknown interaction between CWC22 and RBM8A has also been found which could potentially function in the recruitment of the MAGOH-RBM8A heterodimer. Another interesting link revealed by the C* spliceosome structure was found between EIF4A3 and the EFTUD2. The functional effects of mutating these sites could again be tested in complementation assays with mutated constructs. The Boehm *et al.* (2018) study has identified many targets that need the deposition of the EJC for correct splicing and would serve as a readout of EJC assembly.

In Boehm *et al.* (2018) it was shown that the EJC can influence splicing via two distinct mechanisms: as a binding site for the PSAP complex and by physically masking cryptic splice sites. The splicing order of splice junctions within a transcript does not necessarily occur linearly in 5'-3' direction. Cryptic splice sites can arise at newly formed exon-exon junctions and have the potential to be used in re-splicing. EJCs that are deposited upstream of exon-exon junctions can prevent the usage of these cryptic splice sites and thus maintain correct splicing. The process in which an intron is spliced in several steps is known as recursive splicing and is a productive feature of some long introns (Sibley *et al.*, 2015). In contrast, the re-splicing that occurs when the EJC is depleted leads to aberrant transcripts. The human transcriptome contains many of such splice sites that are not used when the EJC together with the PSAP is present. Potentially, these cryptic splice sites have formed at these positions because there is no selective pressure against them as these regions are normally protected by the EJC.

While a physical block by the EJC explains the protected 3' splice sites, the mechanism behind the splicing regulatory function of RNPS1 has not yet been identified. Tethering SAP18, PNN and RNPS1 individually promoted exon inclusion. A PSAP-binding deficient mutant of RNPS1 was still functional, suggesting that SAP18 and PNN recruit RNPS1 in the tethering assay which is the key mediator of the exon defining function of the PSAP. Interestingly, PSAP-dependent splicing events do not occur upon the depletion of ACIN1, the component of the alternatively formed ASAP complex. Therefore, the PSAP and ASAP complexes regulate different subsets of transcripts (Wang *et al.*, 2018). More work is needed to identify the interaction between the

EJC and the ASAP/PSAP complexes. The RRM domain of RNPS1 is required to form an ASAP/PSAP complex and to interact with EIF4A3 (Boehm *et al.*, 2018). Since neither RNPS1 nor SAP18 were found to directly interact with any of the EJC core components, it is either PNN that links the PSAP complex to the EJC, or the interaction needs a composite binding site of the PSAP complex. ACIN1 on the other hand has been shown to directly interact with a complex formed of EIF4A3, MAGOH and RBM8A (Wang *et al.*, 2018).

Many EJC-dependent targets identified in Boehm *et al.* (2018) were previously described in an earlier study of EJC-dependent splicing and the splicing defects were also detected upon CASC3 depletion (Wang *et al.*, 2014b). We could not replicate these results when performing an siRNA-mediated knockdown of CASC3 (Gerbracht *et al.*, 2019). Even in CASC3 knockout cells these targets were normally spliced (Gerbracht *et al.*, 2019). This leads to the question at which stage of EJC assembly CASC3 joins the EJC. Potentially CASC3 is assembled into the EJC during splicing but is dispensable for the splicing regulatory effect the EJC exerts. Information about the composition of the EJC in the spliceosome comes from *in vitro* splicing reactions followed by either mass-spectrometry analysis or cryo-EM. In the C and C* cryo-EM spliceosome structures the EJC proteins are resolved at 8-15 Å and 4-5 Å with the crystal structure being modeled into the electron density map (Zhang *et al.*, 2017, Zhan *et al.*, 2018). This stands in contrast to mass spectrometry-based analyses of the C spliceosome in which EIF4A3, MAGOH and RBM8A were detected, whereas CASC3 was not (Bessonov *et al.*, 2008, Agafonov *et al.*, 2011). Because CASC3 contains many intrinsically disordered regions, only two fragments are present in the crystal structure of the EJC (Andersen *et al.*, 2006, Bono *et al.*, 2006). Only one of those fragments, the 170-194 stretch which interacts with the RecA2 domain of EIF4A3, is modeled into the spliceosome structures. The space where the more C-terminally resolved fragment of CASC3 would interact with the RecA1 domain of EIF4A3 is occupied by EFTUD2 in the C and C* structures. This discrepancy between the position of CASC3 in the spliceosome structure and in the tetrameric EJC structure indicates that CASC3 can likely only fully bind the EJC after splicing has been completed. Because of the low resolution of the electron density maps, it cannot be excluded that the measured density at this position may not originate from CASC3 but another factor binding the RecA2 domain.

A model in which CASC3 binds to the trimeric EJC core primarily after splicing is supported by several lines of evidence. Firstly, in *in vitro* splicing experiments EIF4A3, MAGOH and RBM8A require splicing to assemble into the EJC, unlike CASC3 (Gehring *et al.*, 2009a). Secondly, the

amount of CASC3 in the cell is limited compared to EIF4A3. RBM8A and MAGOH (Schaab *et al.*, 2012, Singh *et al.*, 2012). Therefore, stochastically not every EJC can be loaded with CASC3. Finally, a recent publication has shown that RNPS1 and CASC3 do not exist within the same EJC (Mabin *et al.*, 2018). RNA-protein immunoprecipitation data has suggested that RNPS1-containing EJCs are bound to rather nuclear located transcripts, whereas CASC3-containing EJCs are associated with predominantly cytoplasmic transcripts. Although CASC3 can potentially shuttle between the cytoplasm and the nucleus it remains to be shown whether it carries out any function in the nucleus.

3.2.2 Cytoplasmic EJC: Nonsense-mediated mRNA decay

RNPS1, CASC3 and UPF3B have been shown to be able to promote NMD in tethering assays (Gehring *et al.*, 2003, Gehring *et al.*, 2005, Boehm *et al.*, 2014). Considering the recent evidence, RNPS1 probably cannot exist together with CASC3 and UPF3B in the same EJC (Mabin *et al.*, 2018). While RNPS1 interacts with the EJC in the nucleus and is essential for its splicing regulatory function, it is not yet known whether it is exported as part of the EJC into the cytoplasm where it could activate NMD. Individual endogenous targets have been reported to be degraded by NMD in an RNPS1-dependent manner (Mabin *et al.*, 2018). Our transcriptome-wide analysis however showed that the stabilization of PTC-containing transcripts in RNPS1 knockdown was marginal (Gerbracht *et al.*, 2019, Figure 4F).

We have observed that CASC3 activates NMD via two distinct routes: as part of the EJC and via its N-terminus (Gerbracht *et al.*, 2019). Several possibilities for the underlying mechanism exist. Since CASC3 has been shown to increase the stability of the EJC when it binds, it could promote NMD by preserving NMD-triggering EJCs. The stabilized EJC would then have an increased likelihood of recruiting NMD factors during each round of translation. An interesting result from the mass spectrometry-based analysis of the EJC composition was that UPF3B is slightly less enriched in the EJCs of CASC3 knockout cells when compared to wild type cells (Gerbracht *et al.*, 2019). Because UPF3B interacts with the EJC via a trimeric binding site formed by EIF4A3, MAGOH and RBM8A, CASC3 may not directly contribute to UPF3B binding. A stabilizing effect on the EJC could explain why more UPF3B was pulled down in the wild type samples. Alternatively, CASC3 could also be responsible to recruit NMD factors, thereby establishing a second route besides UPF3B. To address the question of the relative participations of CASC3 and UPF3B, tethering assays could be used. Tethering CASC3 and co-

depleting UPF3A and UPF3B would reveal whether UPF3B is the factor acting downstream in the NMD activating pathway.

Further studies would be required to disentangle the functions CASC3 has as part of the EJC in contrast to potential effects of its N- and C-termini. For this, a deletion construct consisting only of the SELOR domain could be helpful. The SELOR domain is sufficient to be assembled into the EJC and exert a stabilizing effect (Ballut *et al.*, 2005). A construct which contains the SELOR domain together with a downstream located nuclear export signal is able to fully rescue the NMD deficiency (Gerbracht *et al.*, 2019). Since SELOR localizes into the nucleus, it would have to associate with the EJC there and be exported to the cytoplasm. It would be interesting to explore the role of the nuclear export signal regarding CASC3's ability to promote NMD.

Next, the NMD-inducing effect of CASC3's isolated N-terminus should be further investigated. Tethered full-length CASC3 has been shown to induce NMD in a translation-dependent manner (Boehm *et al.*, 2016). This experiment should be repeated with the isolated CASC3 N-terminus. The depletion of NMD factors such as UPF2 or UPF3B could help to determine whether the N-terminus recruits the known NMD machinery or only selected branches of NMD are activated. Potential interaction partners could be identified by expressing the N-terminus and performing co-immunoprecipitation followed by mass spectrometry.

CASC3's function in NMD could also be linked to the temporal dynamics of transcript degradation. Transcripts that are directly translated after mRNA export and efficiently degraded by NMD perhaps would not be bound by CASC3. Transcripts that are less efficiently degraded or potentially escape NMD could have an increased likelihood of harboring EJCs containing CASC3. Upon subsequent translation, CASC3 would then increase the possibility of successful degradation. Live-cell imaging of NMD transcripts as performed in Hoek *et al.* (2019) in CASC3 knockout cells compared to wild type cells could show whether CASC3 generally activates NMD. This would mean that degradation should occur after more cycles of translation in CASC3 knockout cells compared to wild type cells. If CASC3 acts as a failsafe mechanism to increase the degradation of transcripts that were not yet targeted by the NMD machinery, then the percentage of mRNAs escaping NMD should be lower in wild type cells than in CASC3 knockout cells. A third option is that CASC3 is only needed for the NMD of certain transcripts with distinct features. To address this, endogenous NMD targets that are affected by the CASC3 knockout could be analyzed regarding certain sequence elements or

features such as the position of the PTC in the transcript. The feasibility of such an approach has been shown previously by studying the effect of predictive features of NMD on transcript abundance (Lindeboom *et al.*, 2016, Lindeboom *et al.*, 2019).

Since CASC3 was reported to be able to stimulate translation, this could potentially enhance its stimulatory effect on NMD. The CASC3 knockout cells would be a good resource to study translation efficiency. If CASC3 is a general enhancer of translation, ribosome profiling analysis should show fewer ribosomes engaged with transcripts. On the other hand, it is possible that CASC3 stimulates the translation of selected mRNAs. A ribo-seq analysis would show if individual transcripts are less associated with the translation machinery.

3.3 Pitfalls of CRISPR-Cas9-mediated generation of knockout cell lines

Knocking out non-essential genes by CRISPR-Cas9 genome editing has the exciting promise of studying the complete absence of a gene product. The advantage to performing siRNA-mediated knockdowns is that no remaining protein can obscure functional effects.

The Cas9 nuclease introduces double stranded breaks at the desired locus guided by an RNA. Non-homologous end joining (NHEJ) is a naturally occurring DNA repair mechanism of the cell which does not need a template. Instead, variable insertion or deletions (indels) are introduced at the locus. The introduction of indels at the beginning of an open reading frame result in a frameshift and/or the introduction of premature termination codons. Thus, during translation of this modified mRNA no functional protein can be produced, and PTC-containing mRNAs would be degraded by NMD. This approach has been used in Gerbracht *et al.* (2019) to establish the CASC3 knockout cells. Validating the cell lines by Sanger sequencing of the targeted genomic locus and detection of CASC3 protein using two different antibodies showed that no full-length CASC3 protein could be produced. The presence of a faint protein band on the western blots suggested however that a truncated CASC3 protein was present in two of the three established cell lines. It was confirmed by mass-spectrometry that the cell line H produces low levels of CASC3 and that this gene product is able to interact with the EJC component EIF4A3. Treating the cell line H with siRNAs targeting CASC3 increased the effects observed on the NMD machinery, indicating that the truncated CASC3 protein is functional.

Recently a study investigating an array of 13 knockout cell lines has systematically addressed the issue of residual protein production following gene knockout (Tuladhar *et al.*, 2019). It was

shown that in approximately one third of cell lines the introduction of indels results in aberrant splicing of the mRNA. In approximately one third of cell lines tested, remaining full-length or truncated proteins could be identified. Not all were caused by alternative splicing but instead also resulted from alternative translation initiation sites. A larger analysis of 192 knockout cell lines containing frameshift mutations confirmed the estimate that one third of the cell lines produce residual protein (Smits *et al.*, 2019).

In Gerbracht *et al.* (2019) we could not detect any alternatively spliced transcripts in the CASC3 knockout cell lines. The protein present in the cell line H most likely lacks an N-terminal region because it was detected with antibodies recognizing the central part and its outermost C-terminus. Thus, it seems likely that the truncated protein is produced from an alternative translation start site, although no canonical AUG start codon is in frame downstream of the targeted locus.

When cell lines produce residual proteins or truncated variants, this could mask effects that would be observed in a complete knockout. The occurrence of aberrant splicing or alternative translation start sites could potentially also lead to the production of a protein with novel or dominant negative functions. Care has to be taken when designing and validating CRISPR experiments. The confirmation of protein levels should be carried out by using at least two antibodies targeting different regions of the protein. Guide RNAs are often targeted to the beginning of an open reading frame to insert the frameshift as early as possible. Simultaneously, this increases the chance of a downstream translation initiation event being able to produce functional protein. Two guide RNAs targeting different parts of a gene could be used to increase the chances of a successful knockout. Using two guide RNAs it is also possible to generate clones that lack the entire sequence between them. This was exemplified using the gene BUB1 for which a full knockout has not been feasible before (Raaijmakers & Medema, 2019). A disadvantage of deleting whole genes is that regulatory sequences contained within might be deleted, thereby leading to unpredictable off-target effects.

3.4 Applications and potential enhancements of the xrRNA method

In Boehm *et al.* (2016) and Gerbracht *et al.* (2017) the analysis of steady-state levels of unstable mRNAs using xrRNA-based reporters has led to new biological insights. The relative participation of 5’-3’ decay on degradation induced by AU-rich elements, miRNAs and

premature termination codons was shown, highlighting the broad applicability of the method (Boehm *et al.*, 2016). The analysis also provided the basis to identify new endocleavage sites of the cytokine mRNAs TNF-α and interleukin-6. Also, the effect of the transfection method on the readout of NMD assays was studied using the xrRNA method (Gerbracht *et al.*, 2017). The observation that transiently expressed NMD-sensitive reporters are immune to degradation in certain cell types has an impact for researchers wishing to perform similar experiments. Finally, in Gerbracht *et al.* (2019) reporter mRNAs containing xrRNAs where used to study their degradation by NMD in CASC3 knockout cells.

In the *flavivirus* genus heterogeneous xrRNAs exist that vary in their sequence and composition of XRN1-resistant structures in their 3' UTR (Slonchak & Khromykh, 2018). While the XRN1-resistance of many of these xrRNAs have been tested *in vitro*, it is not known whether using a different set of xrRNAs would improve the detection of XRN1-resistant intermediates in human cells. The xrRNA pseudoknot structures are commonly followed by other types of secondary structures downstream in the 3' UTR of viruses (Clarke *et al.*, 2015). If and how these structures could contribute to XRN1-resistance in reporter assays has not yet been tested. We have only used xrRNA-mRNA-reporters in human cells so far. Since sfRNA production also occurs in other flaviviral hosts such as mosquitos, it is probable that the mechanism of XRN1-inhibition is not limited to human XRN1 (Slonchak & Khromykh, 2018). XRN1's active site which is affected by the xrRNA is highly conserved from arthropods to vertebrates (Jones *et al.*, 2013). Thus, the method could be applied in other model organisms to study mRNA turnover.

A high number of XRN1-resistant RNAs in a host cell was proposed to potentially titrate XRN1 away by acting as a sponge (Charley & Wilusz, 2016). Furthermore, it was shown that sfRNAs of Zika and Dengue viruses bind to proteins involved in the host's post-transcriptional gene regulation (Michalski *et al.*, 2019). However, in our reporter assays we could not detect a depletion of XRN1 which would have led to 3' endocleavage fragments in NMD reporters. Either the amount of xrRNA carrying mRNAs is too little to act as XRN1-sponges, or sequences besides the pseudoknots add to the ability of sfRNA so sequester away proteins.

Blocking exonucleolytic decay from the 3' end would help to completely unravel the exonucleolytic contributions to the degradation of a transcript. Inhibiting 3'-5' decay is

hampered by the fact that the RNA exosome is a multimeric protein complex, employing many different nucleases and that additional exoribonucleases such as DIS3L2 exist.

In Voigt *et al.* (2019) the xrRNA method is described in a protocol format with the second half focusing on using xrRNAs in single-molecule imaging. An issue that was addressed using this approach by Horvathova *et al.* (2017) was the localization of mRNA turnover in the cell. These experiments have shown that the siRNA-induced degradation of a reporter takes place across the cytoplasm and is not enriched in P-bodies. Furthermore, xrRNAs were instrumental in understanding the dynamics of transcripts moving to P-bodies and stress granules during the integrated stress response (Wilbertz *et al.*, 2019). To detect single mRNAs in FISH or live cell imaging, the reporter RNAs contain large arrays of PP7 and MS2 stem loops bound by coat proteins fused to distinct fluorescent proteins. Concerns were raised in yeast regarding the fact that 24 repeats of MS2 binding site bound by MS2-tagged proteins inhibit XRN1-meadiated degradation (Garcia & Parker, 2015, Garcia & Parker, 2016). Fluorogenic aptamers within reporter mRNAs could be a potential solution for this issue. These RNA aptamers bind to a fluorogenic dye, thereby forming a fluorescent complex (Bouhedda *et al.*, 2017). While this approach needs to be improved before being able to be applied for single-molecule imaging, substituting the dye for a fluorogenic protein has already been shown to be applicable (Wu *et al.*, 2019a).

4 REFERENCES

Agafonov DE, Deckert J, Wolf E, Odenwalder P, Bessonov S, Will CL, Urlaub H, Luhrmann R (2011) Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol Cell Biol* 31: 2667-82

Ahn SH, Kim M, Buratowski S (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* 13: 67-76

Akhtar J, Kreim N, Marini F, Mohana G, Brune D, Binder H, Roignant JY (2019) Promoter-proximal pausing mediated by the exon junction complex regulates splicing. *Nat Commun* 10: 521

Akichika S, Hirano S, Shichino Y, Suzuki T, Nishimasu H, Ishitani R, Sugita A, Hirose Y, Iwasaki S, Nureki O, Suzuki T (2019) Cap-specific terminal N (6)-methylation of RNA by an RNA polymerase II-associated methyltransferase. *Science* 363: eaav0080

Akiyama BM, Laurence HM, Massey AR, Costantino DA, Xie X, Yang Y, Shi PY, Nix JC, Beckham JD, Kieft JS (2016) Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science* 354: 1148-1152

Alexandrov A, Colognori D, Shu MD, Steitz JA (2012) Human spliceosomal protein CWC22 plays a role in coupling splicing to exon junction complex deposition and nonsense-mediated decay. *Proc Natl Acad Sci U S A* 109: 21313-8

Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A (2004) A faux 3'-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* 432: 112-8

Andersen CB, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CL, Pedersen JS, Seraphin B, Le Hir H, Andersen GR (2006) Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. *Science* 313: 1968-72

Anderson P (2008) Post-transcriptional control of cytokine production. *Nat Immunol* 9: 353-9

Ashton-Beaucage D, Udell CM, Lavoie H, Baril C, Lefrancois M, Chagnon P, Gendron P, Caron-Lizotte O, Bonneil E, Thibault P, Therrien M (2010) The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in Drosophila. *Cell* 143: 251-62

Atzorn V, Fragapane P, Kiss T (2004) U17/snR30 is a ubiquitous snoRNA with two conserved sequence motifs essential for 18S rRNA production. *Mol Cell Biol* 24: 1769-78

Ballut L, Marchadier B, Baguet A, Tomasetto C, Seraphin B, Le Hir H (2005) The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol* 12: 861-9

Barbosa I, Haque N, Fiorini F, Barrandon C, Tomasetto C, Blanchette M, Le Hir H (2012) Human CWC22 escorts the helicase eIF4AIII to spliceosomes and promotes exon junction complex assembly. *Nat Struct Mol Biol* 19: 983-90

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136: 215-33

Bensaude O (2011) Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* 2: 103-108

Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM (1998) Localization of ASH1 mRNA particles in living yeast. *Mol Cell* 2: 437-45

Bessonov S, Anokhina M, Krasauskas A, Golas MM, Sander B, Will CL, Urlaub H, Stark H, Luhrmann R (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* 16: 2384-403

Bessonov S, Anokhina M, Will CL, Urlaub H, Luhrmann R (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* 452: 846-50

Boehm V, Britto-Borges T, Steckelberg AL, Singh KK, Gerbracht JV, Gueney E, Blazquez L, Altmuller J, Dieterich C, Gehring NH (2018) Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. *Mol Cell* 72: 482-495 e7

Boehm V, Gerbracht JV, Marx MC, Gehring NH (2016) Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences. *Nat Commun* 7: 13691

Boehm V, Haberman N, Ottens F, Ule J, Gehring NH (2014) 3' UTR length and messenger ribonucleoprotein composition determine endocleavage efficiencies at termination codons. *Cell Rep* 9: 555-68

Bokar JA, Shambaugh ME, Polayes D, Matera AG, Rottman FM (1997) Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA* 3: 1233-47

Bono F, Cook AG, Grunwald M, Ebert J, Conti E (2010) Nuclear import mechanism of the EJC component Mago-Y14 revealed by structural studies of importin 13. *Mol Cell* 37: 211-22

Bono F, Ebert J, Lorentzen E, Conti E (2006) The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell* 126: 713-25

Bouhedda F, Autour A, Ryckelynck M (2017) Light-Up RNA Aptamers and Their Cognate Fluorogens: From Their Development to Their Applications. *Int J Mol Sci* 19

Boulias K, Toczydlowska-Socha D, Hawley BR, Liberman N, Takashima K, Zaccara S, Guez T, Vasseur JJ, Debart F, Aravind L, Jaffrey SR, Greer EL (2019) Identification of the m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of m(6)Am in the Transcriptome. *Mol Cell* 75: 631-643 e8

Braun JE, Truffault V, Boland A, Huntzinger E, Chang CT, Haas G, Weichenrieder O, Coles M, Izaurralde E (2012) A direct interaction between DCP1 and XRN1 couples mRNA decapping to 5' exonucleolytic degradation. *Nat Struct Mol Biol* 19: 1324-31

Brendza RP, Serbus LR, Duffy JB, Saxton WM (2000) A function for kinesin I in the posterior transport of oskar mRNA and Staufen protein. *Science* 289: 2120-2

Buckley SM, Aranda-Orgilles B, Strikoudis A, Apostolou E, Loizou E, Moran-Crusio K, Farnsworth CL, Koller AA, Dasgupta R, Silva JC, Stadtfeld M, Hochedlinger K, Chen EI, Aifantis I (2012) Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. *Cell Stem Cell* 11: 783-98

Buratowski S (2003) The CTD code. *Nat Struct Biol* 10: 679-80

Burger K, Muhl B, Kellner M, Rohrmoser M, Gruber-Eber A, Windhager L, Friedel CC, Dolken L, Eick D (2013) 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol* 10: 1623-30

Bursac S, Brdovcak MC, Donati G, Volarevic S (2014) Activation of the tumor suppressor p53 upon impairment of ribosome biogenesis. *Biochim Biophys Acta* 1842: 817-30

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 55: 611-22

Cech TR (2009) Evolution of biological catalysis: ribozyme to RNP enzyme. *Cold Spring Harb Symp Quant Biol* 74: 11-6

Chan CC, Dostie J, Diem MD, Feng W, Mann M, Rappsilber J, Dreyfuss G (2004) eIF4A3 is a novel component of the exon junction complex. *RNA* 10: 200-9

Chan WK, Huang L, Gudikote JP, Chang YF, Imam JS, MacLean JA, 2nd, Wilkinson MF (2007)

An alternative branch of the nonsense-mediated decay pathway. *EMBO J* 26: 1820-30

Chang CT, Bercovich N, Loh B, Jonas S, Izaurralde E (2014a) The activation of the decapping enzyme DCP2 by DCP1 occurs on the EDC4 scaffold and involves a conserved loop in DCP1. *Nucleic Acids Res* 42: 5217-33

Chang CT, Muthukumar S, Weber R, Levdansky Y, Chen Y, Bhandari D, Igreja C, Wohlbold L, Valkov E, Izaurralde E (2019) A low-complexity region in human XRN1 directly recruits deadenylation and decapping factors in 5'-3' messenger RNA decay. *Nucleic Acids Res* 47: 9282-9295

Chang H, Lim J, Ha M, Kim VN (2014b) TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. *Mol Cell* 53: 1044-52

Chapman EG, Costantino DA, Rabe JL, Moon SL, Wilusz J, Nix JC, Kieft JS (2014) The structural basis of pathogenic subgenomic flavivirus RNA (sfRNA) production. *Science* 344: 307-10

Charley PA, Wilusz J (2016) Standing your ground to exoribonucleases: Function of Flavivirus long non-coding RNAs. *Virus Res* 212: 70-7

Chazal PE, Daguenet E, Wendling C, Ulryck N, Tomasetto C, Sargueil B, Le Hir H (2013) EJC core component MLN51 interacts with eIF3 and activates translation. *Proc Natl Acad Sci U S A* 110: 5903-8

Chen CY, Shyu AB (2011) Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip Rev RNA* 2: 167-83

Cheng H, Dufu K, Lee CS, Hsu JL, Dias A, Reed R (2006) Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* 127: 1389-400

Cheng Y, Kiess AP, Herman JM, Pomper MG, Meltzer SJ, Abraham JM (2015) Phosphorus-32, a clinically available drug, inhibits cancer growth by inducing DNA double-strand breakage. *PLoS One* 10: e0128152

Choe J, Ryu I, Park OH, Park J, Cho H, Yoo JS, Chi SW, Kim MK, Song HK, Kim YK (2014) eIF4AIII enhances translation of nuclear cap-binding complex-bound mRNAs by promoting disruption of secondary structures in 5'UTR. *Proc Natl Acad Sci U S A* 111: E4577-86

Chong HY, Leow CY, Abdul Majeed AB, Leow CH (2019) Flavivirus infection-A review of immunopathogenesis, immunological response, and immunodiagnosis. *Virus Res* 274: 197770

Ciganda M, Williams N (2011) Eukaryotic 5S rRNA biogenesis. *Wiley Interdiscip Rev RNA* 2: 523-33

Ciriello S (2014) Separate functions of BTZ during post--transcriptional gene regulation. In Mathematisch-Naturwissenschaftliche Fakultät, Universität zu Köln

Clarke BD, Roby JA, Slonchak A, Khromykh AA (2015) Functional non-coding RNAs derived from the flavivirus 3' untranslated region. *Virus Res* 206: 53-61

Collart MA (2016) The Ccr4-Not complex is a key regulator of eukaryotic gene expression. *Wiley Interdiscip Rev RNA* 7: 438-54

Colombo M, Karousis ED, Bourquin J, Bruggmann R, Muhlemann O (2017) Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways. *RNA* 23: 189-201

Core L, Adelman K (2019) Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev* 33: 960-982

Cortazar MA, Sheridan RM, Erickson B, Fong N, Glover-Cutter K, Brannan K, Bentley DL (2019) Control of RNA Pol II Speed by PNUTS-PP1 and Spt5 Dephosphorylation Facilitates Termination by a "Sitting Duck Torpedo" Mechanism. *Mol Cell* 76: 896-908 e4

Coulon A, Ferguson ML, de Turris V, Palangat M, Chow CC, Larson DR (2014) Kinetic competition

during the transcription cycle results in stochastic RNA processing. *Elife* 3

Cramer P (2019a) Eukaryotic Transcription Turns 50. *Cell* 179: 808-812

Cramer P (2019b) Organization and regulation of gene transcription. *Nature* 573: 45-54

D'Orazio KN, Wu CC, Sinha N, Loll-Krippleber R, Brown GW, Green R (2019) The endonuclease Cue2 cleaves mRNAs at stalled ribosomes during No Go Decay. *Elife* 8

Das S, Singer RH, Yoon YJ (2019) The travels of mRNAs in neurons: do they know where they are going? *Curr Opin Neurobiol* 57: 110-116

De Conti L, Baralle M, Buratti E (2013) Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA* 4: 49-60

Degot S, Regnier CH, Wendling C, Chenard MP, Rio MC, Tomasetto C (2002) Metastatic Lymph Node 51, a novel nucleo-cytoplasmic protein overexpressed in breast cancer. *Oncogene* 21: 4422-34

Diem MD, Chan CC, Younis I, Dreyfuss G (2007) PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat Struct Mol Biol* 14: 1173-9

Dostie J, Dreyfuss G (2002) Translation is required to remove Y14 from mRNAs in the cytoplasm. *Curr Biol* 12: 1060-7

Dujardin G, Lafaille C, de la Mata M, Marasco LE, Munoz MJ, Le Jossic-Corcos C, Corcos L, Kornblihtt AR (2014) How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell* 54: 683-90

Durand S, Lykke-Andersen J (2013) Nonsense-mediated mRNA decay occurs during eIF4F-dependent translation in human cells. *Nat Struct Mol Biol* 20: 702-9

Eberle AB, Lykke-Andersen S, Muhlemann O, Jensen TH (2009) SMG6 promotes endonucleolytic cleavage of nonsense mRNA in human cells. *Nat Struct Mol Biol* 16: 49-55

Eberle AB, Stalder L, Mathys H, Orozco RZ, Muhlemann O (2008) Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region. *PLoS Biol* 6: e92

Eulalio A, Behm-Ansmant I, Schweizer D, Izaurralde E (2007) P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol* 27: 3970-81

Fan XC, Steitz JA (1998) Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *EMBO J* 17: 3448-60

Ferraiuolo MA, Lee CS, Ler LW, Hsu JL, Costa-Mattioli M, Luo MJ, Reed R, Sonenberg N (2004) A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. *Proc Natl Acad Sci U S A* 101: 4118-23

Fica SM, Oubridge C, Wilkinson ME, Newman AJ, Nagai K (2019) A human postcatalytic spliceosome structure reveals essential roles of metazoan factors for exon ligation. *Science* 363: 710-714

Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, Diener K, Jones K, Fu XD, Bentley DL (2014) Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev* 28: 2663-76

Frith MC, Pheasant M, Mattick JS (2005) The amazing complexity of the human transcriptome. *Eur J Hum Genet* 13: 894-7

Funk A, Truong K, Nagasaki T, Torres S, Floden N, Balmori Melian E, Edmonds J, Dong H, Shi PY, Khromykh AA (2010) RNA structures required for production of subgenomic flavivirus RNA. *J Virol* 84: 11407-17

Garcia-Maurino SM, Rivero-Rodriguez F, Velazquez-Cruz A, Hernandez-Vellisca M, Diaz-Quintana A, De la Rosa MA, Diaz-Moreno I (2017) RNA Binding Protein Regulation and Cross-Talk in the Control of AU-rich mRNA Fate. *Front Mol Biosci* 4: 71

Garcia JF, Parker R (2015) MS2 coat proteins bound to yeast mRNAs block 5' to 3'

degradation and trap mRNA decay products: implications for the localization of mRNAs by MS2-MCP system. *RNA* 21: 1393-5

Garcia JF, Parker R (2016) Ubiquitous accumulation of 3' mRNA decay fragments in Saccharomyces cerevisiae mRNAs with chromosomally integrated MS2 arrays. *RNA* 22: 657-9

Gehring NH, Kunz JB, Neu-Yilik G, Breit S, Viegas MH, Hentze MW, Kulozik AE (2005) Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements. *Mol Cell* 20: 65-75

Gehring NH, Lamprinaki S, Hentze MW, Kulozik AE (2009a) The hierarchy of exon-junction complex assembly by the spliceosome explains key features of mammalian nonsense-mediated mRNA decay. *PLoS Biol* 7: e1000120

Gehring NH, Lamprinaki S, Kulozik AE, Hentze MW (2009b) Disassembly of exon junction complexes by PYM. *Cell* 137: 536-48

Gehring NH, Neu-Yilik G, Schell T, Hentze MW, Kulozik AE (2003) Y14 and hUpf3b form an NMD-activating complex. *Mol Cell* 11: 939-49

Gehring NH, Wahle E, Fischer U (2017) Deciphering the mRNP Code: RNA-Bound Determinants of Post-Transcriptional Gene Regulation. *Trends Biochem Sci* 42: 369-382

Gerbracht JV, Boehm V, Britto-Borges T, Kallabis S, Wiederstein JL, Ciriello S, Aschemeier DU, Krüger M, Frese CK, Altmüller J, Dieterich C, Gehring NH (2019) CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex. *bioRxiv*: 811018

Gerbracht JV, Boehm V, Gehring NH (2017) Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells. *Sci Rep* 7: 10616

Gerbracht JV, Gehring NH (2018) The exon junction complex: structural insights into a faithful companion of mammalian mRNPs. *Biochem Soc Trans* 46: 153-161

Ghosh A, Lima CD (2010) Enzymology of RNA cap synthesis. *Wiley Interdiscip Rev RNA* 1: 152-72

Ghosh S, Marchand V, Gaspar I, Ephrussi A (2012) Control of RNP motility and localization by a splicing-dependent structure in oskar mRNA. *Nat Struct Mol Biol* 19: 441-9

Ghosh S, Obrdlik A, Marchand V, Ephrussi A (2014) The EJC binding and dissociating activity of PYM is regulated in Drosophila. *PLoS Genet* 10: e1004455

Glock C, Heumuller M, Schuman EM (2017) mRNA transport & local translation in neurons. *Curr Opin Neurobiol* 45: 169-177

Golomb L, Volarevic S, Oren M (2014) p53 and ribosome biogenesis stress: the essentials. *FEBS Lett* 588: 2571-9

Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* 89: 5547-51

Gossen M, Freundlieb S, Bender G, Muller G, Hillen W, Bujard H (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science* 268: 1766-9

Graveley BR, Hertel KJ, Maniatis T (2001) The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA* 7: 806-18

Gray NK, Hentze MW (1994) Iron regulatory protein prevents binding of the 43S translation pre-initiation complex to ferritin and eALAS mRNAs. *EMBO J* 13: 3882-91

Gromadzka AM, Steckelberg AL, Singh KK, Hofmann K, Gehring NH (2016) A short conserved motif in ALYREF directs cap- and EJC-dependent assembly of export complexes on spliced mRNAs. *Nucleic Acids Res* 44: 2348-61

Ha M, Kim VN (2014) Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* 15: 509-24

Hachet O, Ephrussi A (2001) Drosophila Y14 shuttles to the posterior of the oocyte and is

required for oskar mRNA transport. *Curr Biol* 11: 1666-74

Harlen KM, Churchman LS (2017) The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* 18: 263-273

Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, Urlaub H, Kastner B, Luhrmann R, Stark H (2018) Structure and Conformational Dynamics of the Human Spliceosomal B(act) Complex. *Cell* 172: 454-464 e11

Hayashi R, Handler D, Ish-Horowicz D, Brennecke J (2014) The exon junction complex is required for definition and excision of neighboring introns in Drosophila. *Genes Dev* 28: 1772-85

Heath CG, Viphakone N, Wilson SA (2016) The role of TREX in gene expression and disease. *Biochem J* 473: 2911-35

Henras AK, Plisson-Chastang C, O'Donohue MF, Chakraborty A, Gleizes PE (2015) An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip Rev RNA* 6: 225-42

Hentze MW, Caughman SW, Rouault TA, Barriocanal JG, Dancis A, Harford JB, Klausner RD (1987) Identification of the iron-responsive element for the translational regulation of human ferritin mRNA. *Science* 238: 1570-3

Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, von Haeseler A, Zuber J, Ameres SL (2017) Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* 14: 1198-1204

Hoek TA, Khuperkar D, Lindeboom RGH, Sonneveld S, Verhagen BMP, Boersma S, Vermeulen M, Tanenbaum ME (2019) Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. *Mol Cell* 75: 324-339 e11

Horvathova I, Voigt F, Kotrys AV, Zhan Y, Artus-Revel CG, Eglinger J, Stadler MB, Giorgetti L, Chao JA (2017) The Dynamics of mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells. *Mol Cell* 68: 615-625 e9

Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat Rev Genet* 12: 99-110

Huntzinger E, Kashima I, Fauser M, Sauliere J, Izaurralde E (2008) SMG6 is the catalytic endonuclease that cleaves mRNAs containing nonsense codons in metazoan. *RNA* 14: 2609-17

Ivanov A, Mikhailova T, Eliseev B, Yeramala L, Sokolova E, Susorov D, Shuvalov A, Schaffitzel C, Alkalaeva E (2016) PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Res* 44: 7766-76

Jones CI, Grima DP, Waldron JA, Jones S, Parker HN, Newbury SF (2013) The 5'-3' exoribonuclease Pacman (Xrn1) regulates expression of the heat shock protein Hsp67Bc and the microRNA miR-277-3p in Drosophila wing imaginal discs. *RNA Biol* 10: 1345-55

Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, Ohno M, Dreyfuss G, Ohno S (2006) Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes Dev* 20: 355-67

Kastner B, Will CL, Stark H, Luhrmann R (2019) Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb Perspect Biol* 11: a032417

Kataoka N, Yong J, Kim VN, Velazquez F, Perkinson RA, Wang F, Dreyfuss G (2000) Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell* 6: 673-82

Ke S, Pandya-Jones A, Saito Y, Fak JJ, Vagbo CB, Geula S, Hanna JH, Black DL, Darnell JE, Jr., Darnell RB (2017) m(6)A mRNA modifications are deposited in nascent pre-mRNA and are not required for splicing but do specify cytoplasmic turnover. *Genes Dev* 31: 990-1006

Khatter H, Myasnikov AG, Natchiar SK, Klaholz BP (2015) Structure of the human 80S ribosome. *Nature* 520: 640-5

Kieft JS, Rabe JL, Chapman EG (2015) New hypotheses derived from the structure of a flaviviral Xrn1-resistant RNA: Conservation, folding, and host adaptation. *RNA Biol* 12: 1169-77

Kiss AM, Jady BE, Bertrand E, Kiss T (2004) Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol* 24: 5797-807

Komarnitsky P, Cho EJ, Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* 14: 2452-60

Koyuncu S, Saez I, Lee HJ, Gutierrez-Garcia R, Pokrzywa W, Fatima A, Hoppe T, Vilchez D (2018) The ubiquitin ligase UBR5 suppresses proteostasis collapse in pluripotent stem cells from Huntington's disease patients. *Nat Commun* 9: 2886

Kramer S (2017) Simultaneous detection of mRNA transcription and decay intermediates by dual colour single mRNA FISH on subcellular resolution. *Nucleic Acids Res* 45: e49

Kressler D, Hurt E, Bassler J (2017) A Puzzle of Life: Crafting Ribosomal Subunits. *Trends Biochem Sci* 42: 640-654

Kumar A, Clerici M, Muckenfuss LM, Passmore LA, Jinek M (2019) Mechanistic insights into mRNA 3'-end processing. *Curr Opin Struct Biol* 59: 143-150

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446: 926-9

Le Hir H, Gatfield D, Braun IC, Forler D, Izaurralde E (2001a) The protein Mago provides a link between splicing and mRNA localization. *EMBO Rep* 2: 1119-24

Le Hir H, Gatfield D, Izaurralde E, Moore MJ (2001b) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J* 20: 4987-97

Le Hir H, Izaurralde E, Maquat LE, Moore MJ (2000) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* 19: 6860-9

Le Hir H, Nott A, Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* 28: 215-20

Lejeune F, Ishigaki Y, Li X, Maquat LE (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J* 21: 3536-45

Leppek K, Schott J, Reitter S, Poetz F, Hammond MC, Stoecklin G (2013) Roquin promotes constitutive mRNA decay via a conserved class of stem-loop recognition motifs. *Cell* 153: 869-81

Lim J, Ha M, Chang H, Kwon SC, Simanshu DK, Patel DJ, Kim VN (2014) Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* 159: 1365-76

Lindeboom RG, Supek F, Lehner B (2016) The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet* 48: 1112-8

Lindeboom RGH, Vermeulen M, Lehner B, Supek F (2019) The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat Genet* 51: 1645-1651

Linder P, Jankowsky E (2011) From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* 12: 505-16

Loh B, Jonas S, Izaurralde E (2013) The SMG5-SMG7 heterodimer directly recruits the CCR4-NOT deadenylase complex to mRNAs containing nonsense codons via interaction with POP2. *Genes Dev* 27: 2125-38

Lu S, Cullen BR (2003) Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA* 9: 618-30

Luo Y, Na Z, Slavoff SA (2018) P-Bodies: Composition, Properties, and Functions. *Biochemistry* 57: 2424-2431

Lykke-Andersen S, Chen Y, Ardal BR, Lilje B, Waage J, Sandelin A, Jensen TH (2014) Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev* 28: 2498-517

Ma XM, Yoon SO, Richardson CJ, Julich K, Blenis J (2008) SKAR links pre-mRNA splicing to mTOR/S6K1-mediated enhanced translation efficiency of spliced mRNAs. *Cell* 133: 303-13

Mabin JW, Woodward LA, Patton RD, Yi Z, Jia M, Wysocki VH, Bundschuh R, Singh G (2018) The Exon Junction Complex Undergoes a Compositional Switch that Alters mRNP Structure and Nonsense-Mediated mRNA Decay Activity. *Cell Rep* 25: 2431-2446 e7

Malone CD, Mestdagh C, Akhtar J, Kreim N, Deinhard P, Sachidanandam R, Treisman J, Roignant JY (2014) The exon junction complex controls transposable element activity by ensuring faithful splicing of the piwi transcript. *Genes Dev* 28: 1786-99

Mao H, Pilaz LJ, McMahon JJ, Golzio C, Wu D, Shi L, Katsanis N, Silver DL (2015) Rbm8a haploinsufficiency disrupts embryonic cortical development resulting in microcephaly. *J Neurosci* 35: 7003-18

Mattick JS (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep* 2: 986-91

Mauer J, Luo X, Blanjoie A, Jiao X, Grozhik AV, Patil DP, Linder B, Pickering BF, Vasseur JJ, Chen Q, Gross SS, Elemento O, Debart F, Kiledjian M, Jaffrey SR (2017) Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature* 541: 371-375

Mauer J, Sindelar M, Despic V, Guez T, Hawley BR, Vasseur JJ, Rentmeister A, Gross SS, Pellizzoni L, Debart F, Goodarzi H, Jaffrey SR (2019) FTO controls reversible m(6)Am RNA methylation during snRNA biogenesis. *Nat Chem Biol* 15: 340-347

McCracken S, Fong N, Rosonina E, Yankulov K, Brothers G, Siderovski D, Hessel A, Foster S, Shuman S, Bentley DL (1997) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* 11: 3306-18

McStay B (2016) Nucleolar organizer regions: genomic 'dark matter' requiring illumination. *Genes Dev* 30: 1598-610

Meyer KD, Patil DP, Zhou J, Zinoviev A, Skabkin MA, Elemento O, Pestova TV, Qian SB, Jaffrey SR (2015) 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* 163: 999-1010

Michalski D, Ontiveros JG, Russo J, Charley PA, Anderson JR, Heck AM, Geiss BJ, Wilusz J (2019) Zika virus noncoding sfRNAs sequester multiple host-derived RNA-binding proteins and modulate mRNA decay and splicing during infection. *J Biol Chem* 294: 16282-16296

Mino T, Murakawa Y, Fukao A, Vandenbon A, Wessels HH, Ori D, Uehata T, Tartey S, Akira S, Suzuki Y, Vinuesa CG, Ohler U, Standley DM, Landthaler M, Fujiwara T, Takeuchi O (2015) Regnase-1 and Roquin Regulate a Common Element in Inflammatory mRNAs by Spatiotemporally Distinct Mechanisms. *Cell* 161: 1058-1073

Mitchell P, Petfalski E, Shevchenko A, Mann M, Tollervey D (1997) The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'-->5' exoribonucleases. *Cell* 91: 457-66

Mohr SE, Dillon ST, Boswell RE (2001) The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize oskar mRNA during Drosophila oogenesis. *Genes Dev* 15: 2886-99

Muller-McNicoll M, Botti V, de Jesus Domingues AM, Brandl H, Schwich OD, Steiner MC, Curk T, Poser I, Zarnack K, Neugebauer KM (2016) SR

proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes Dev* 30: 553-66

Mullner EW, Kuhn LC (1988) A stem-loop in the 3' untranslated region mediates iron-dependent regulation of transferrin receptor mRNA stability in the cytoplasm. *Cell* 53: 815-25

Mullner EW, Neupert B, Kuhn LC (1989) A specific mRNA binding factor regulates the iron-dependent stability of cytoplasmic transferrin receptor mRNA. *Cell* 58: 373-82

Murachelli AG, Ebert J, Basquin C, Le Hir H, Conti E (2012) The structure of the ASAP core complex reveals the existence of a Pinin-containing PSAP complex. *Nat Struct Mol Biol* 19: 378-86

Nagarajan VK, Jones CI, Newbury SF, Green PJ (2013) XRN 5'-->3' exoribonucleases: structure, mechanisms and functions. *Biochim Biophys Acta* 1829: 590-603

Nagy E, Maquat LE (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 23: 198-9

Nasif S, Contu L, Muhlemann O (2018) Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression. *Semin Cell Dev Biol* 75: 78-87

Neu-Yilik G, Gehring NH, Thermann R, Frede U, Hentze MW, Kulozik AE (2001) Splicing and 3' end formation in the definition of nonsense-mediated decay-competent human beta-globin mRNPs. *EMBO J* 20: 532-40

Neu-Yilik G, Raimondeau E, Eliseev B, Yeramala L, Amthor B, Deniaud A, Huard K, Kerschgens K, Hentze MW, Schaffitzel C, Kulozik AE (2017) Dual function of UPF3B in early and late translation termination. *EMBO J* 36: 2968-2986

Newmark PA, Boswell RE (1994) The mago nashi locus encodes an essential product required for germ plasm assembly in Drosophila. *Development* 120: 1303-13

Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M, Jr. (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev* 21: 708-18

Nielsen KH, Chamieh H, Andersen CB, Fredslund F, Hamborg K, Le Hir H, Andersen GR (2009) Mechanism of ATP turnover inhibition in the EJC. *RNA* 15: 67-75

Noble CG, Song H (2007) MLN51 stimulates the RNA-helicase activity of eIF4AIII. *PLoS One* 2: e303

Nojima T, Rebelo K, Gomes T, Grosso AR, Proudfoot NJ, Carmo-Fonseca M (2018) RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol Cell* 72: 369-379 e4

Nott A, Le Hir H, Moore MJ (2004) Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev* 18: 210-22

Nott A, Meislin SH, Moore MJ (2003) A quantitative analysis of intron effects on mammalian gene expression. *RNA* 9: 607-17

Okada-Katsuhata Y, Yamashita A, Kutsuzawa K, Izumi N, Hirahara F, Ohno S (2012) N- and C-terminal Upf1 phosphorylations create binding platforms for SMG-6 and SMG-5:SMG-7 during NMD. *Nucleic Acids Res* 40: 1251-66

Pabis M, Neufeld N, Steiner MC, Bojic T, Shav-Tal Y, Neugebauer KM (2013) The nuclear cap-binding complex interacts with the U4/U6.U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA* 19: 1054-63

Palacios IM, Gatfield D, St Johnston D, Izaurralde E (2004) An eIF4AIII-containing complex required for mRNA localization and nonsense-mediated mRNA decay. *Nature* 427: 753-7

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40: 1413-5

Pantopoulos K (2004) Iron metabolism and the IRE/IRP regulatory system: an update. *Ann N Y Acad Sci* 1012: 1-13

Papasaikas P, Valcarcel J (2016) The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci* 41: 33-45

Peck SA, Hughes KD, Victorino JF, Mosley AL (2019) Writing a wrong: Coupled RNA polymerase II transcription and RNA quality control. *Wiley Interdiscip Rev RNA* 10: e1529

Peixeiro I, Inacio A, Barbosa C, Silva AL, Liebhaber SA, Romao L (2012) Interaction of PABPC1 with the translation initiation complex is critical to the NMD resistance of AUG-proximal nonsense mutations. *Nucleic Acids Res* 40: 1160-73

Pena C, Hurt E, Panse VG (2017) Eukaryotic ribosome assembly, transport and quality control. *Nat Struct Mol Biol* 24: 689-699

Pijlman GP, Funk A, Kondratieva N, Leung J, Torres S, van der Aa L, Liu WJ, Palmenberg AC, Shi PY, Hall RA, Khromykh AA (2008) A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe* 4: 579-91

Piovesan A, Caracausi M, Antonaros F, Pelleri MC, Vitale L (2016) GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database (Oxford)* 2016

Pokrywka NJ, Stephenson EC (1995) Microtubules are a general component of mRNA localization systems in Drosophila oocytes. *Dev Biol* 167: 363-70

Porrua O, Libri D (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* 16: 190-202

Raaijmakers JA, Medema RH (2019) Killing a zombie: a full deletion of the BUB1 gene in HAP1 cells. *EMBO J* 38: e102423

Ramanathan A, Robb GB, Chan SH (2016) mRNA capping: biological functions and applications. *Nucleic Acids Res* 44: 7511-26

Rennel E, Gerwins P (2002) How to make tetracycline-regulated transgene expression go on and off. *Analytical biochemistry* 309: 79-84

Robberson BL, Cote GJ, Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 10: 84-94

Roignant JY, Treisman JE (2010) Exon junction complex subunits are required to splice Drosophila MAP kinase, a large heterochromatic gene. *Cell* 143: 238-50

Rufener SC, Muhlemann O (2013) eIF4E-bound mRNPs are substrates for nonsense-mediated mRNA decay in mammalian cells. *Nat Struct Mol Biol* 20: 710-7

Ryu I, Kim YK (2017) Translation initiation mediated by nuclear cap-binding protein complex. *BMB Rep* 50: 186-193

Saez I, Gerbracht JV, Koyuncu S, Lee HJ, Horn M, Kroef V, Denzel MS, Dieterich C, Gehring NH, Vilchez D (2020) The E3 ubiquitin ligase UBR5 interacts with the H/ACA ribonucleoprotein complex and regulates ribosomal RNA biogenesis in embryonic stem cells. *FEBS Lett* 594: 175-188

Saez I, Koyuncu S, Gutierrez-Garcia R, Dieterich C, Vilchez D (2018) Insights into the ubiquitin-proteasome system of human embryonic stem cells. *Sci Rep* 8: 4092

Saldi T, Cortazar MA, Sheridan RM, Bentley DL (2016) Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol* 428: 2623-2635

Sanduja S, Blanco FF, Dixon DA (2011) The roles of TTP and BRF proteins in regulated mRNA decay. *Wiley Interdiscip Rev RNA* 2: 42-57

Sato H, Maquat LE (2009) Remodeling of the pioneer translation initiation complex involves translation and the karyopherin importin beta. *Genes Dev* 23: 2537-50

Sauliere J, Murigneux V, Wang Z, Marquenet E, Barbosa I, Le Tonqueze O, Audic Y, Paillard L, Roest Crollius H, Le Hir H (2012) CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. *Nat Struct Mol Biol* 19: 1124-31

Schaab C, Geiger T, Stoehr G, Cox J, Mann M (2012) Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol Cell Proteomics* 11: M111 014068

Schmid M, Jensen TH (2018) Controlling nuclear RNA levels. *Nat Rev Genet* 19: 518-529

Schmidt SA, Foley PL, Jeong DH, Rymarquis LA, Doyle F, Tenenbaum SA, Belasco JG, Green PJ (2015) Identification of SMG6 cleavage sites and a preferred RNA cleavage motif by global analysis of endogenous NMD targets in human cells. *Nucleic Acids Res* 43: 309-23

Schneider M, Will CL, Anokhina M, Tazi J, Urlaub H, Luhrmann R (2010) Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol Cell* 38: 223-35

Schnettler E, Sterken MG, Leung JY, Metz SW, Geertsema C, Goldbach RW, Vlak JM, Kohl A, Khromykh AA, Pijlman GP (2012) Noncoding flavivirus RNA displays RNA interference suppressor activity in insect and Mammalian cells. *J Virol* 86: 13486-500

Schoenberg DR (2011) Mechanisms of endonuclease-mediated mRNA decay. *Wiley Interdiscip Rev RNA* 2: 582-600

Schoenberg DR, Maquat LE (2012) Regulation of cytoplasmic mRNA decay. *Nat Rev Genet* 13: 246-59

Schroeder SC, Schwer B, Shuman S, Bentley D (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* 14: 2435-40

Schuessler A, Funk A, Lazear HM, Cooper DA, Torres S, Daffis S, Jha BK, Kumagai Y, Takeuchi O, Hertzog P, Silverman R, Akira S, Barton DJ, Diamond MS, Khromykh AA (2012) West Nile virus noncoding subgenomic RNA contributes to viral evasion of the type I interferon-mediated antiviral response. *J Virol* 86: 5708-18

Schuller AP, Green R (2018) Roadblocks and resolutions in eukaryotic translation. *Nat Rev Mol Cell Biol* 19: 526-541

Schwerk C, Prasad J, Degenhardt K, Erdjument-Bromage H, White E, Tempst P, Kidd VJ, Manley JL, Lahti JM, Reinberg D (2003) ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol Cell Biol* 23: 2981-90

Shearer RF, Iconomou M, Watts CK, Saunders DN (2015) Functional Roles of the E3 Ubiquitin Ligase UBR5 in Cancer. *Mol Cancer Res* 13: 1523-32

Shibuya T, Tange TO, Sonenberg N, Moore MJ (2004) eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* 11: 346-51

Shoemaker CJ, Green R (2012) Translation drives mRNA quality control. *Nat Struct Mol Biol* 19: 594-601

Sibley CR, Emmett W, Blazquez L, Faro A, Haberman N, Briese M, Trabzuni D, Ryten M, Weale ME, Hardy J, Modic M, Curk T, Wilson SW, Plagnol V, Ule J (2015) Recursive splicing in long vertebrate genes. *Nature* 521: 371-375

Silva PA, Pereira CF, Dalebout TJ, Spaan WJ, Bredenbeek PJ (2010) An RNA pseudoknot is required for production of yellow fever virus subgenomic RNA by the host nuclease XRN1. *J Virol* 84: 11395-406

Silver DL, Watkins-Chow DE, Schreck KC, Pierfelice TJ, Larson DM, Burnetti AJ, Liaw HJ, Myung K, Walsh CA, Gaiano N, Pavan WJ (2010) The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat Neurosci* 13: 551-8

Simms CL, Thomas EN, Zaher HS (2017) Ribosome-based quality control of mRNA and nascent peptides. *Wiley Interdiscip Rev RNA* 8: e1366

Simon B, Masiewicz P, Ephrussi A, Carlomagno T (2015) The structure of the SOLE element of oskar mRNA. *RNA* 21: 1444-53

Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ (2012) The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* 151: 750-764

Singh G, Rebbapragada I, Lykke-Andersen J (2008) A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol* 6: e111

Sloan KE, Warda AS, Sharma S, Entian KD, Lafontaine DLJ, Bohnsack MT (2017) Tuning the ribosome: The influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol* 14: 1138-1152

Slonchak A, Khromykh AA (2018) Subgenomic flaviviral RNAs: What do we know after the first decade of research. *Antiviral Res* 159: 13-25

Smits AH, Ziebell F, Joberty G, Zinn N, Mueller WF, Clauder-Munster S, Eberhard D, Falth Savitski M, Grandi P, Jakob P, Michon AM, Sun H, Tessmer K, Burckstummer T, Bantscheff M, Steinmetz LM, Drewes G, Huber W (2019) Biological plasticity rescues target activity in CRISPR knock outs. *Nat Methods* 16: 1087-1093

So BR, Di C, Cai Z, Venters CC, Guo J, Oh JM, Arai C, Dreyfuss G (2019) A Complex of U1 snRNP with Cleavage and Polyadenylation Factors Controls Telescripting, Regulating mRNA Transcription in Human Cells. *Mol Cell* 76: 590-599 e4

Sommer S, Lavi U, Darnell JE, Jr. (1978) The absolute frequency of labeled N-6-methyladenosine in HeLa cell messenger RNA decreases with label time. *J Mol Biol* 124: 487-99

Steckelberg AL, Altmueller J, Dieterich C, Gehring NH (2015) CWC22-dependent pre-mRNA splicing and eIF4A3 binding enables global deposition of exon junction complexes. *Nucleic Acids Res* 43: 4687-700

Steckelberg AL, Boehm V, Gromadzka AM, Gehring NH (2012) CWC22 connects pre-mRNA splicing and exon junction complex assembly. *Cell Rep* 2: 454-61

Steckelberg AL, Gehring NH (2014) Studying the composition of mRNPs in vitro using splicing-competent cell extracts. *Methods* 65: 342-9

Strasser K, Masuda S, Mason P, Pfannstiel J, Oppizzi M, Rodriguez-Navarro S, Rondon AG, Aguilera A, Struhl K, Reed R, Hurt E (2002) TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* 417: 304-8

Tafforeau L, Zorbas C, Langhendries JL, Mullineux ST, Stamatopoulou V, Mullier R, Wacheul L, Lafontaine DL (2013) The complexity of human ribosome biogenesis revealed by systematic nucleolar screening of Pre-rRNA processing factors. *Mol Cell* 51: 539-51

Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159: 635-46

Tange TO, Shibuya T, Jurica MS, Moore MJ (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA* 11: 1869-83

Thomas MP, Liu X, Whangbo J, McCrossan G, Sanborn KB, Basar E, Walch M, Lieberman J (2015) Apoptosis Triggers Specific, Rapid, and Global mRNA Decay with 3' Uridylated Intermediates Degraded by DIS3L2. *Cell Rep* 11: 1079-89

Toma KG, Rebbapragada I, Durand S, Lykke-Andersen J (2015) Identification of elements in human long 3' UTRs that inhibit nonsense-mediated decay. *RNA* 21: 887-97

Trcek T, Larson DR, Moldon A, Query CC, Singer RH (2011) Single-molecule mRNA decay measurements reveal promoter- regulated mRNA stability in yeast. *Cell* 147: 1484-97

Trcek T, Sato H, Singer RH, Maquat LE (2013) Temporal and spatial characterization of nonsense-mediated mRNA decay. *Genes Dev* 27: 541-51

Tuladhar R, Yeu Y, Tyler Piazza J, Tan Z, Rene Clemenceau J, Wu X, Barrett Q, Herbert J, Mathews DH, Kim J, Hyun Hwang T, Lum L (2019) CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat Commun* 10: 4056

Uchida Y, Chiba T, Kurimoto R, Asahara H (2019) Post-transcriptional regulation of inflammation by RNA-binding proteins via cis-elements of mRNAs. *J Biochem* 166: 375-382

Ule J, Blencowe BJ (2019) Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* 76: 329-345

Valencia P, Dias AP, Reed R (2008) Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc Natl Acad Sci U S A* 105: 3386-91

van Eeden FJ, Palacios IM, Petronczki M, Weston MJ, St Johnston D (2001) Barentsz is essential for the posterior localization of oskar mRNA and colocalizes with it to the posterior pole. *J Cell Biol* 154: 511-23

Viphakone N, Sudbery I, Griffith L, Heath CG, Sims D, Wilson SA (2019) Co-transcriptional Loading of RNA Export Factors Shapes the Human Transcriptome. *Mol Cell* 75: 310-323 e8

Voigt F, Gerbracht JV, Boehm V, Horvathova I, Eglinger J, Chao JA, Gehring NH (2019) Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts. *Nature protocols* 14: 1603-1633

Wada T, Becskei A (2017) Impact of Methods on the Measurement of mRNA Turnover. *Int J Mol Sci* 18

Wang DO, Ninomiya K, Mori C, Koyama A, Haan M, Kitabatake M, Hagiwara M, Chida K, Takahashi SI, Ohno M, Kataoka N (2017) Transport Granules Bound with Nuclear Cap Binding Protein and Exon Junction Complex Are Associated with Microtubules and Spatially Separated from eIF4E Granules and P Bodies in Human Neuronal Processes. *Front Mol Biosci* 4: 93

Wang M, Pestov DG (2011) 5'-end surveillance by Xrn2 acts as a shared mechanism for mammalian pre-rRNA maturation and decay. *Nucleic Acids Res* 39: 1811-22

Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, Ren B, Pan T, He C (2014a) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 505: 117-20

Wang X, Zhao BS, Roundtree IA, Lu Z, Han D, Ma H, Weng X, Chen K, Shi H, He C (2015) N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* 161: 1388-99

Wang Z, Ballut L, Barbosa I, Le Hir H (2018) Exon Junction Complexes can have distinct functional flavours to regulate specific splicing events. *Sci Rep* 8: 9509

Wang Z, Kiledjian M (2001) Functional link between the mammalian exosome and mRNA decapping. *Cell* 107: 751-62

Wang Z, Murigneux V, Le Hir H (2014b) Transcriptome-wide modulation of splicing by the exon junction complex. *Genome Biol* 15: 551

Warner JR (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24: 437-40

Watkins NJ, Bohnsack MT (2012) The box C/D and H/ACA snoRNPs: key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* 3: 397-414

Webster MW, Chen YH, Stowell JAW, Alhusaini N, Sweet T, Graveley BR, Coller J, Passmore LA (2018) mRNA Deadenylation Is Coupled to Translation Rates by the Differential Activities of Ccr4-Not Nucleases. *Mol Cell* 70: 1089-1100 e8

Weil TT (2014) mRNA localization in the Drosophila germline. *RNA Biol* 11: 1010-8

Wiegand HL, Lu S, Cullen BR (2003) Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. *Proc Natl Acad Sci U S A* 100: 11327-32

Wilbertz JH, Voigt F, Horvathova I, Roth G, Zhan Y, Chao JA (2019) Single-Molecule Imaging of mRNA Localization and Regulation during the Integrated Stress Response. *Mol Cell* 73: 946-958 e7

Will CL, Luhrmann R (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3

Wilson T, Treisman R (1988) Removal of poly(A) and consequent degradation of c-fos mRNA facilitated by 3' AU-rich sequences. *Nature* 336: 396-9

Wolf J, Passmore LA (2014) mRNA deadenylation by Pan2-Pan3. *Biochem Soc Trans* 42: 184-7

Wu J, Zaccara S, Khuperkar D, Kim H, Tanenbaum ME, Jaffrey SR (2019a) Live imaging of mRNA using RNA-stabilized fluorogenic proteins. *Nat Methods* 16: 862-865

Wu Q, Medina SG, Kushawah G, DeVore ML, Castellano LA, Hand JM, Wright M, Bazzini AA (2019b) Translation affects mRNA stability in a codon-dependent manner in human cells. *Elife* 8

Yamashita A, Chang TC, Yamashita Y, Zhu W, Zhong Z, Chen CY, Shyu AB (2005a) Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat Struct Mol Biol* 12: 1054-63

Yamashita A, Kashima I, Ohno S (2005b) The role of SMG-1 in nonsense-mediated mRNA decay. *Biochim Biophys Acta* 1754: 305-15

Yan X, Hoek TA, Vale RD, Tanenbaum ME (2016) Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* 165: 976-89

Yi H, Park J, Ha M, Lim J, Chang H, Kim VN (2018) PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay. *Mol Cell* 70: 1081-1088 e5

Zhan X, Yan C, Zhang X, Lei J, Shi Y (2018) Structure of a human catalytic step I spliceosome. *Science* 359: 537-545

Zhang X, Yan C, Hang J, Finci LI, Lei J, Shi Y (2017) An Atomic Structure of the Human Spliceosome. *Cell* 169: 918-929 e14

Zhang X, Yan C, Zhan X, Li L, Lei J, Shi Y (2018) Structure of the human activated spliceosome in three conformational states. *Cell Res* 28: 307-322

Zheng G, Dahl JA, Niu Y, Fedorcsak P, Huang CM, Li CJ, Vagbo CB, Shi Y, Wang WL, Song SH, Lu Z, Bosmans RP, Dai Q, Hao YJ, Yang X, Zhao WM, Tong WM, Wang XJ, Bogdan F, Furu K et al. (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* 49: 18-29

Zhu J, Mayeda A, Krainer AR (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 8: 1351-61

Zund D, Gruber AR, Zavolan M, Muhlemann O (2013) Translation-dependent displacement of UPF1 from coding sequences causes its enrichment in 3' UTRs. *Nat Struct Mol Biol* 20: 936-43

5 SUMMARY

The output of gene products needs to be regulated in order for cells to function in development and react to signals or stress. As a first step of gene expression, genes are transcribed into RNA. Messenger RNAs (mRNAs) are produced which carry the genetic information from the DNA to ribosomes, the sites of protein biosynthesis in the cell. Also other RNA species are transcribed that perform regulatory or catalytical functions within the gene expression machinery. All RNAs in the cell undergo processing and quality control as soon as they are transcribed. Furthermore, they are bound by RNA binding proteins, forming messenger ribonucleoproteins (mRNPs). This process of post-transcriptional gene regulation ensures that the produced RNA species are stable and localize to the correct subcellular department. In this cumulative thesis three distinct aspects of post-transcriptional gene regulation are investigated.

In the first part the processing of ribosomal RNAs (rRNAs) is analyzed. As the most abundant RNA species in the cell, mature rRNAs are derived from a polycistronic precursor RNA that undergoes cleavage and modification before being exported into the cytoplasm with ribosomal proteins, thus forming mature ribosomal subunits. A myriad of processing factors has been described which are involved in rRNA maturation. Here, a link between the ubiquitin E3 ligase UBR5 and the ribosomal processing machinery is identified in human embryonic stem cells. UBR5 interacts with components of the box H/ACA snoRNP which carries out pseudouridylation modifications of rRNAs. It is shown that depletion of UBR5 in stem cells impairs rRNA maturation and ribosomal biosynthesis as a whole.

The second part of this thesis is focused on the exon junction complex (EJC) which is an essential RNA-binding complex in metazoan cells. The EJC is deposited during splicing upstream of exon-exon junctions of mRNAs and stays bound until it is removed during translation. During the lifetime of an mRNA, the EJC fulfills critical functions in splicing, mRNA export and translation. It is also important for the mRNA surveillance pathway nonsense-mediated mRNA decay (NMD). The assembly of the individual EJC components is intimately linked to the splicing process, however the exact dynamics and interactions at play are not fully understood. The recently published structures of the human spliceosome have given insights into the positioning of EJCs components within spliceosome intermediates. These advances are discussed in the thesis and conclusions for further studies are drawn. Next, a

new function of the EJC in connection with splicing is described: after its deposition, the EJC together with associated splicing factors inhibits the usage of potential splice sites in the transcriptome, thereby maintaining splicing fidelity. Since this function is only observed for three of the four described core EJC components, the molecular function of the fourth EJC component CASC3 is investigated. Using CRISPR-Cas9-generated knockout cells it is shown that CASC3 is important for the function of the EJC in the NMD pathway.

Finally, in the last part of the thesis the process of mRNA turnover is described. All mRNAs eventually undergo degradation in order to maintain dynamics in gene expression. Furthermore, faulty mRNAs are targeted by quality control pathways which leads to their degradation. Finally, mRNAs such as cytokine transcripts exist that need to be tightly controlled and are therefore unstable. The analysis of mRNA turnover is challenging, since different nucleolytic pathways can act on the degradation of transcripts. Here, a novel method is presented that employs viral structures which are resistant to the human 5'-3' exonuclease XRN1. By using these structures in unstable reporter transcripts, the participation of exo- and endonucleolytic degradation can be determined. This approach is also described in a protocol format and can be used to monitor the degradation of mRNAs in single cells using live-cell imaging.

6 ZUSAMMENFASSUNG

Die Produktion von Genprodukten muss reguliert werden, damit Zellen in der Entwicklung funktionieren und auf Signale oder Stress reagieren können. Als erster Schritt der Genexpression werden Gene in RNA transkribiert. Es werden Boten-RNAs (englisch messenger RNAs, mRNAs) hergestellt, die die genetische Information von der DNA zu Ribosomen transportieren. Auch andere RNA-Spezies, die regulatorische oder katalytische Funktionen innerhalb der Genexpressionsmaschine ausführen, werden transkribiert. Alle RNAs in der Zelle werden sofort nach ihrer Transkription einer Prozessierung und Qualitätskontrolle unterzogen. Darüber hinaus interagieren sie mit RNA-bindende Proteinen und bilden Boten-Ribonukleoproteine (englisch messenger ribonucleoproteins, mRNPs). Dieser Prozess der posttranskriptionellen Genregulation stellt sicher, dass die produzierten RNA-Spezies stabil sind und korrekt in der Zelle lokalisieren. In dieser kumulativen Arbeit werden drei verschiedene Aspekte der posttranskriptionellen Genregulation untersucht.

Im ersten Teil wird die Maturierung von ribosomalen RNAs (rRNAs) analysiert. Als die am häufigsten vorkommende RNA-Spezies in der Zelle werden reife rRNAs von einer polycistronischen Vorläufer-RNA abgeleitet. Diese wird gespalten und modifiziert, bevor sie mit ribosomalen Proteinen in das Zytoplasma exportiert wird. Hierdurch entstehen reife ribosomale Untereinheiten. In der Literatur ist eine Vielzahl von Verarbeitungsfaktoren beschrieben, die an der Reifung der rRNA beteiligt sind. Hier wird ein funktioneller Zusammenhang zwischen der Ubiquitin-E3-Ligase UBR5 und der ribosomalen Prozessierungsmaschinerie in humanen embryonalen Stammzellen identifiziert. UBR5 interagiert mit Komponenten des Box H/ACA-snoRNPs, welcher Pseudouridylierungs-modifikationen von rRNAs durchführt. Es wird gezeigt, dass die Depletion von UBR5 in Stammzellen die Reifung der rRNA und die ribosomale Biogenese insgesamt beeinträchtigt.

Der zweite Teil dieser Arbeit befasst sich mit dem Exon Junction Complex (EJC), einem essentiellen RNA-bindenden Komplex in metazoischen Zellen. Der EJC wird während des Spleißens stromaufwärts von Exon-Exon-Grenzen auf mRNAs assembliert und bleibt dort gebunden, bis er während der Translation entfernt wird. Während der Lebensdauer einer mRNA erfüllt der EJC fundamentelle Aufgaben beim Spleißen, mRNA-Export und bei der Translation. Es ist auch wichtig für den mRNA-Qualitätskontrollmechanismus nonsense-vermittelter mRNA-Abbau (NMD). Die Assemblierung der einzelnen EJC-Komponenten ist eng

mit dem Spleißen verbunden, jedoch ist die Dynamik dieses Prozesses nicht vollständig bekannt. Die kürzlich veröffentlichten Strukturen des menschlichen Spleißosoms haben Einblicke in die Positionierung von EJC-Komponenten innerhalb von Spleißosom-intermediaten gegeben. Diese Fortschritte werden in dieser Arbeit diskutiert und Schlussfolgerungen für weitere Studien gezogen. Als nächstes wird eine neu identifizierte Funktion des EJC im Zusammenhang mit Spleißen beschrieben: Nach seiner Positionierung hemmt das EJC zusammen mit assoziierten Spleißfaktoren die Verwendung potentieller Spleißstellen im Transkriptom, wodurch die Spleißgenauigkeit erhalten bleibt. Da diese Funktion nur für drei der vier beschriebenen EJC-Kernkomponenten zutrifft, wird anschließend die molekulare Funktion der vierten EJC-Komponente CASC3 untersucht. Mit CRISPR-Cas9-generierten Knockout-Zellen wird gezeigt, dass CASC3 für die Funktion des EJC im NMD-Signalweg von Bedeutung ist.

Abschließend wird im letzten Teil der Arbeit der Prozess des mRNA-Abbaus beschrieben. Alle mRNAs werden letztendlich degradiert, um die Dynamik der Genexpression aufrechtzuerhalten. Darüber hinaus werden fehlerhafte mRNAs durch Qualitätskontrollwege angegriffen, die zu ihrem Abbau führen. Schließlich gibt es mRNAs wie Cytokin-Transkripte, deren Expression streng kontrolliert werden muss und welche daher instabil sind. Die Analyse des mRNA-Abbaus ist eine Herausforderung, da verschiedene nukleolytische Wege auf den Abbau von Transkripten einwirken. Hier wird eine neue Methode vorgestellt, die auf viralen Strukturen basiert, welche resistent gegen Abbau durch die humane 5’-3’-Exonuklease XRN1 sind. Durch die Verwendung dieser Strukturen in instabilen Reporter-Transkripten kann die Beteiligung des exo- und endonukleolytischen Abbaus bestimmt werden. Dieser Ansatz wird auch in einem Protokoll beschrieben und kann darüber hinaus verwendet werden, um den Abbau von mRNAs in einzelnen Zellen unter Verwendung von Live-Cell-Imaging zu überwachen.

7 AUTHOR CONTRIBUTIONS

- Saez, I, **Gerbracht, JV**, Koyuncu, S, Lee, HJ, Horn, M, Kroef, V, Denzel, MS, Dieterich, C, Gehring, NH and Vilchez, D (2020) The E3 ubiquitin ligase UBR5 interacts with the H/ACA ribonucleoprotein complex and regulates ribosomal RNA biogenesis in embryonic stem cells. FEBS Lett, 594, 175-188.

I.S. and D.V. conceived and supervised the study. Together with N.H.G. I designed, performed and analysed the experiments to assess rRNA maturation (Figure 3). Together with I.S., D.V. and N.H.G I wrote the manuscript.

- **Gerbracht JV**, Gehring NH (2018) The exon junction complex: structural insights into a faithful companion of mammalian mRNPs. Biochem Soc Trans 46: 153-161

Together with N.H.G I wrote the review.

- Boehm V, Britto-Borges T, Steckelberg AL, Singh KK, **Gerbracht JV**, Gueney E, Blazquez L, Altmuller J, Dieterich C, Gehring NH (2018) Exon Junction Complexes Suppress Spurious Splice Sites to Safeguard Transcriptome Integrity. Mol Cell 72: 482-495 e7

N.H.G. and V.B. conceived the study. I designed the HSD17B10 and TUFM minigene reporters (Figure 3 and 4) and supervised E.G. who performed the experiments. Furthermore, I was involved in the analysis and interpretation of the data and wrote the manuscript together with N.H.G, V.B., A.-L.S., T.B.B, C.D. and K.K.S.

- **Gerbracht JV,** Boehm V, Britto-Borges T, Kallabis S, Wiederstein JL, Ciriello S, Aschemeier DU, Krüger M, Frese CK, Altmüller J, Dieterich C, Gehring NH (2019) CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex. bioRxiv: 811018

Together with V.B. and N.H.G. I conceived and designed the study. I performed the majority of experiments and was involved in the analysis, interpretation and visualization of all data. Together with V.B. and N.H.G I wrote the manuscript. Detailed author contributions are listed at the end of the paper

- Boehm V*, **Gerbracht JV***, Marx MC, Gehring NH (2016) Interrogating the degradation pathways of unstable mRNAs with XRN1-resistant sequences. Nat Commun 7: 136

N.H.G. and V.B. conceived the study. I performed the experiments shown in Figures 2, 6, 7 and 8 and was involved in analyzing, interpreting and visualizing the data. Together with N.H.G. and V.B. I wrote the manuscript.

- **Gerbracht JV**, Boehm V, Gehring NH (2017) Plasmid transfection influences the readout of nonsense-mediated mRNA decay reporter assays in human cells. Sci Rep 7: 10616

Together with N.H.G. and V.B. I conceived and designed the study. I performed all the experiments and was involved in the analysis, interpretation and visualization of the data. Together with N.H.G. I wrote the manuscript.

- Voigt F*, **Gerbracht JV***, Boehm V, Horvathova I, Eglinger J, Chao JA, Gehring NH (2019) Detection and quantification of RNA decay intermediates using XRN1-resistant reporter transcripts. Nat Protoc 14: 1603-1633

Together with V.B. I developed the northern blotting and qPCR protocol and generated the respective Figures 3-5. Together with F.V., V.B., I.H., J.A.C. and N.H.G. I wrote the manuscript.

* These authors contributed equally

8 ACKNOWLEDGEMENTS

Foremost, I would like to thank Prof. Niels Gehring for giving me the opportunity to conduct my thesis research in his lab. Many thanks for providing the intriguing projects and for the advice and support during this time.

I am very grateful to Prof. Kay Hofmann, firstly for having been a member of my thesis committee during my time as a doctoral student and secondly for evaluating this work as a second reviewer.

I would also like to thank Prof. Ulrich Baumann for chairing the thesis defense.

The Graduate School of Biological Sciences I would like to thank for the opportunities to improve my skills outside of the research lab, especially as part of the organizing committee of “Crossroads in Biology 2019”.

The current and former members of the Gehring lab, I would like to thank for their helpfulness and the warm and friendly working atmosphere.

The research groups of the third floor of the Institute for Genetics, the Leptin, Riemer and Schnetz labs, I would like to thank for sharing equipment, reagents and expertise. I also value the time we spent together at many cake seminars and carnival events over the years.

I am also very grateful to Dr. Volker Böhm, Lindsay Gerbracht, Sabrina Kückelmann, Lena Schlautmann and Dr. Franziska Ottens for proofreading.

Many thanks also to my family and friends for their love and support.

ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit − einschließlich Tabellen, Karten und Abbildungen −, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie − abgesehen von oben angegebenen Teilpublikationen − noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Niels Gehring betreut worden.

Köln, den _____ Jennifer V. Gerbracht